ASTES

Advances in Science, Technology & Engineering Systems Journal

VOLUME 4-ISSUE 5|SEPT-OCT 2019

www.astesj.com
ISSN: 2415-6698

EDITORIAL BOARD
Editor-in-Chief

Prof. Passerini Kazmerski
University of Chicago, USA

Editorial Board Members

Prof. Rehan Ullah Khan
Qassim University, Saudi Arabia

Prof. María Jesús Espinosa
Universidad Tecnológica Metropolitana, Mexico

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

Dr. Nguyen Tung Linh
Electric Power University, Vietnam

Tariq Kamal
University of Nottingham, UK

Sakarya University, Turkey

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

Mohamed Mohamed Abdel-Daim
Suez Canal University, Egypt

Dr. Omeje Maxwell
Covenant University, Nigeria

Prof. Majida Ali Abed Meshari
Tikrit University Campus, Iraq

Dr. Heba Afify
MTI university, Cairo, Egypt

Regional Editors

Dr. Hung-Wei Wu
Kun Shan University, Taiwan

Dr. Maryam Asghari
Shahid Ashrafi Esfahani, Iran

Dr. Shakir Ali
Aligarh Muslim University, India

Dr. Ahmet Kayabasi
Karamanoglu Mehmetbey University, Turkey

Dr. Ebubekir Altuntas
Gaziosmanpasa University, Turkey

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

Dr. Shagufta Haneef
Aalborg University, Denmark

Dr. Gomathi Periasamy
Mekelle University, Ethiopia

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

Aamir Nawaz
Gomal University, Pakistan

Abdullah El-Bayoumi
Cairo University, Egypt

Ayham Hassan Abazid
Jordan university of science and technology, Jordan

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

Editorial

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

Current Issue features key papers related to multidisciplinary domains involving complex system stemming from numerous disciplines; this is exactly how this journal differs from other interdisciplinary and multidisciplinary engineering journals. This issue contains 53 accepted papers in Computer Science domain.

Editor-in-chief
Prof. Passerini Kazmersk

ADVANCES IN SCIENCE, TECHNOLOGY AND ENGINEERING SYSTEMS JOURNAL

Volume 4 Issue 5 **September-October 2019**

CONTENTS

Logic Error Detection System based on Structure Pattern and Error Degree 01
Yuto Yoshizawa, Yutaka Watanobe

Melanoma detection using color and texture features in computer vision systems 16
Antonio Fuduli, Pierangelo Veltri, Eugenio Vocaturo, Ester Zumpano

Application of Open-Source Optimization Library "Extremum" to the Synthesis of Feedback Control of a Satellite 23
Andrei Panteleev, Valentin Panovskiy

Fully Homomorphic Encryption Scheme Based On Complex Numbers 30
Khalil Hariss, Maroun Chamoun, Abed Ellatif Samhat

Low-Cost and Accurate Computational System for Efficiency Measures over Photovoltaic Arrays 39
Hector L. Nunez-Ramırez, Gloria G. Carvalho-Kassar, Freddy C. Brito-Maestre, Yaremi I. Gamboa-Maldonado, Luis A. Santos-Avendano, Carlelines Gavidia-Toro, Orlando Villarroel-Ramos, Dino Di Rosa-Ulloa

Detecting Malicious Assembly using Convolutional, Recurrent Neural Networks 46
Michael Santacroce, Daniel Koranek, Rashmi Jha

Proposal for a Security Model for a Popular Voting System Process in Latin America 53
Segundo Moisés Toapanta Toapanta, Allan Fabricio German Diaz, Darío Fernando Huilcapi Subia, Luis Enrique Mafla Gallegos

Mechanical Testing Methods for Body-Powered Upper-Limb Prostheses: A Case Study 61
Renato Mio, Midori Sanchez, Quino Valverde, José Lara, Francisco Rumiche

Assessment of the Quality and Sustainability Implications of FIFO and LIFO Inventory Policies through System Dynamics 69
Phoebe Lim Ching, Jose Edgar Mutuc, John Anthony Jose

A Study on Development of Evaluation Metrics for Learners in Physical Computing 82
Woochun Jun

The Effect of Fuel Additives on Pyrorated Biodiesel Blends on the Performance of a Diesel Power Generator 88
Maroa Semakula, Freddie Inambao

Toward a Smart Campus Using IoT: Framework for Safety and Security System on a University Campus 97
Alghamdi Abdullah, Mohammed Thanoon, Anwar Alsulami

pH Neutralization Process Control based on Active Disturbance Rejection Control 104
Mohanad Hamad Eljack Elameen, Mona Bahaeldin Musa Hago, Wen Tan

Forecasting Bio-economic Effects in the Milk Production based on the Potential of Animals for Productivity and Viability 110
Gennadii Georgievich Cherepanov, Anatolii Ivanovich Mikhalskii, Zhanna Anatolievna Novosrltseva

Survey on Semantic Similarity Based on Document Clustering 115
Rowaida Khalil Ibrahim, Subhi Rafeeq Mohammed Zeebaree, Karwan Fahmi Sami Jacksi

Green Inventory Supply Chain Model with Inflation under Permissible Delay in Finite Planning Horizon 123
Seema Saxena, Vikramjeet Singh, Rajesh Kumar Gupta, Nitin Kumar Mishra, Pushpinder Singh

Optimization Method of Wideband Multilayer Meander-Line Polarizer using Semi-Analytical approach and Application to 6-18GHz Polarizer including test with Horn Antenna 132
Wafa Abdouni-Abdallah, Muhammad Saeed Khan, Athanasios Konstantinidis, Anne-Claude Tarot, Aziz Ouacha

Development of Teaching Aids Selection Standards in Maker Education 139
Woochun Jun

Application of Fractal Algorithms to Identify Cardiovascular Diseases in ECG Signals 143
Gustavo Gallegos Dávalos, Pedro Freddy Huamaní Navarrete

Satisfaction of the Graduate for the Continuous Improvement of Educational Quality in UNTELS 151
Omar Freddy Chamorro Atalaya, Teodoro Neri Díaz Leyva, Dora Yvonne Arce Santillan, Jorge Isaac Castro Bedriñana, Denisse Marie Barrientos Pichilingue, Avid Roman-Gonzalez

Android Application to Detect Cat Disease Using an Expert System 158
Alvina Aulia, Yogi Udjaja, Irma Kartika Wairooy, Adithia Panji Hutama, Dinar Khansa Shabira, Sabilal Muhtadin

Chatting with Plants (Orchids) in Automated Smart Farming using IoT, Fuzzy Logic and Chatbot 163
Samruan Wiangsamut, Phatthanaphong Chomphuwiset, Suchart Khummanee

DFIG Defects Diagnosis Method for Wind Energy Conversion Chain 174
Fatima El Hammouchi, Lamia El Menzhi, Abdallah Saad

The Impact of Using Upgraded Biogas on Generator Performance 186
Wilson Babu Musinguzi, Ibrahim Luqman Mpungu

Prediction of Non-Communicable Diseases Using Class Comparison Data Mining 193
Ammar Al-Dallal, Amina Al-Moosa

Integrating Diacritics Restoration and Question Classification into Vietnamese Question Answering System 207
Bui Thanh Hung

Improvement opportunities of a Simulation/Expert System Approach for Manufacturing System Sizing: A review and proposal 213
Lajmi Imen, Masmoudi Wassim, Elleuch Mounir, Chtourou Hedi

Fuzzy Simulation of Historical Associative Thesaurus 224
Olena Basalkevych, Olexandr Basalkevych

Transformation of Solar Energy to Electricity with Photovoltaic Systems, Reduction of Electrical Consumption and Installation of LED Lamps in the ITSAO 234
Jeremias Bravo Tapia, Juan José Rivera Toledo, Francisco Ramos Guzmán, Yoxkin Estévez Martínez

Improve the accuracy of short-term forecasting algorithms by Standardized Load Profile and Support Regression Vector: Case study Vietnam 243
Nguyen Tuan Dung, Nguyen Thanh Phuong

Design and Simulation of a Doppler-Radar RF Front-End Transceiver 250
Mohamed Elhefnawy

CISELexia: Computer-Based Method for Improving Self-Awareness in Children with Dyslexia 258
Aulia Rahmawati, Emil Robert Kaburuan, Anditya Arifianto, Nahda Kurnia Juniati

Performances of Thermoelectric Cooling System for Cooling Flowing Water in a Water Block 268
Mirmanto Mirmanto, Yesung Allo Padang, Zefri Ardian

Proposal Methodology of Planning and Location of Loading/Unloading Spaces for Urban Freight Vehicle: A Case Study 273
Moufad Imane, Jawab Fouad

Sentiment Analysis of Transjakarta Based on Twitter using Convolutional Neural Network 281
Kevin Yudi, Suharjito

An Integrated Framework for Pronominal Anaphora Resolution in Malayalam 287
Ajees Arimbassery Pareed, Sumam Mary Idicula

Spectrum Requirements for Control and Non-payload Communication of Maritime Autonomous Surface Ship 294
Ho Namgung

Resonator Influence Simulation of Designed Close-Open Standing Wave Thermoacoustic Engine 300
Sugiyanto, Samsul Kamal, Joko Waluyo, Adhika Widyaparaga

Priority Incorporated Zone Based Distributed Clustering Algorithm for Heterogeneous Wireless Sensor Network 306
Hamid Ali Abed AL-Asadi, Abdulhadi Alhassani, Nor Azura Ahmed Hambali, Mustafa Abdulazeez AlSibahee, Saif Ali Alwazzan, Ali Mohammed Jasim

Stability Analysis of a Linear Model Predictive Control and its Application in a Water Recovery Process 314
Paolo Mercorelli

Estimating Academic results from Trainees' Activities in Programming Exercises Using Four Types of Machine Learning 321
Toshiyasu Kato, Yuki Terawaki, Yasushi Kodama, Teruhiko Unoki, Yasushi Kambayashi

Improving the Performance of Hadoop Framework Using Optimization Process in the Information Management 327
Ramachandran Ravi Sowmiyasree, Nachimuthu Maheswari, Manickam Sivagami

A Modular Multipurpose, Parameter Centered Electronic Health Record Architecture 334
Norbert Maggi, Luca Douglas Magnoni, Carmelina Ruggiero, Roberta Gazzarata, Mauro Giacomini

Direction of Arrival Using 2-D Matrix Pencil Method 341
Mohammed Amine ihedrane, Adiba EL FAdl, Seddik Bri

Determination of Submerged Breakwater Location for Coastal Protection in Panjang Island Waters Jepara 346
Denny Nugroho Sugianto, Purwanto, Gentur Handoyo, Aulia Oktaviani, Kunarso, Muhammad Zainuri, Agus Indarjo

Analysis of Information Security for a Voting Process for Sectional Governments in Ecuador 352
Segundo Moisés Toapanta Toapanta, Steven Xavier Romo Sañicela, Danny Wilfrido Barona Valencia, Luis Enrique Mafla Gallegos

LED Lighting Implementation as a Strategy for Energy Saving in Energetically Sustainable Campus 360
Pablo Pérez-Gosende

Verification of the Usefulness of Personal Authentication with Aerial Input Numerals Using Leap Motion 369
Shun Yamamoto, Momoyo Ito, Shin-ichi Ito, Minoru Fukumi

Artificial Bee Colony-Optimized LSTM for Bitcoin Price Prediction 375
Andary Dadang Yuliyono, Abba Suganda Girsang

Modelling and Control of The PV Power System Dedicated to The Greenhouse for Reduced Energy Consumption 384
Manel Jomaa ur-Laper, Mohamed Yassine Allani ur-Laper, Abdelkader Mami ur-Laper

Transmission Line Restoration Using ERS Structure 394
Sushri Mukherjee, Sumana Chattaraj, Md. Irfan Khan, Dharmbir Prasad, Pradip Barua, Harish Agarwal

EKMC: Ensemble of kNN using MetaCost for Efficient Anomaly Detection 401
Niranjan A, Akshobhya K M, P Deepa Shenoy, Venugopal K R

A Big Data Security Layer Meta-Model Proposition 409
Allae Erraissi, Abdessamad Belangour

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 5, 1-15 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Logic Error Detection System based on Structure Pattern and Error Degree

Yuto Yoshizawa*, Yutaka Watanobe

The University of Aizu, Computer Science and Engineering, Fukushima 965-8580, Japan

ARTICLE INFO

Article history:
Received: 12 May, 2019
Accepted: 14 August, 2019
Online: 06 September, 2019

Keywords:
Logic Error Detection
Programming Education
e-Learning
Online Judge

ABSTRACT

The importance of programming skills has increased with advances in information and communication technology (ICT). However, the difficulty of learning programming is a major problem for novices. Therefore, we propose a logic error detection algorithm based on structure patterns, which are an index of similarity based on abstract syntax trees, and error degree, which is a measure of appropriateness for feedback. We define structure patterns and error degree and present the proposed algorithm. In addition, we develop a Logic Error Detector (LED) Application Programming Interface (API) based on the proposed algorithm. An implementation of the proposed algorithm is used in experiments using actual data from an e-learning system. The results show that the proposed algorithm can accurately detect logic errors in many programs solving problems in the Introduction to Programming set.

1 Introduction

The Fourth Industrial Revolution has increased the demand for engineers [1] [2] [3]. Until recently, computer science was learned by university students, often in engineering fields. Because of a shortage of engineers in many countries, attempts have been undertaken to introduce computer science into secondary school education.

A study suggested that the high school ICT curriculum should focus on computer science [4]. A report by the Royal Society of the United Kingdom emphasized the importance of computer science skills in primary education [5]. The report pointed out problems in ICT education, such as those in class content and the loss of opportunities for students. In addition, it proposed to redefine ICT into three areas, namely computer science, information technology, and digital literacy, among which computer science is a particularly important area. The movement to implement programming education is accelerating globally.

For introducing programming in school education, many problems have been identified and various teaching methods have been proposed [6] [7]. A report summarizing three years of compulsory programming education in the UK published in 2014 found that it was patchy and fragile [8]. The report briefly discussed the difficulty of programming education.

A student's inherent interest in computer programming affects learning. Learners with an interest in programming can effectively learn in large groups with a single teacher, as is done in the university. Such learners can solve problems by investigating them or asking questions. However, traditional teaching methods may be unsuitable for novices and learners who are not interested in programming. One of the most effective ways for novices to learn is one-on-one learning with a teacher. This is especially true for programming because programs that accomplish a given task can be written in various ways. Teachers must thus strictly evaluate the program written by a learner.

Programming includes complex language constructs and requires logical thinking, similar to that used in mathematics. Programmers must consider algorithms, memory space, execution time, and code size. However, exhaustive and quantitative evaluation of programs by a human teacher is difficult. In addition, programs may not work as expecting due to various bugs. It is difficult even for experienced teachers to deal with every bug.

To improve this situation, it is necessary to assign a knowledgeable teacher to several novice programmers. However, this is not possible in classes that introduce programming due to a lack of human resources. To solve this problem, methods for estimating the difficulty of programming problems using cluster analysis and problem recommendation algorithms have been proposed [9] [10]. In addition, attempts have been made to identify early crisis situations experienced by novice programmers using clustering [11]. However, many studies considered only optimizing human resources and

*Corresponding Author: Yuto Yoshizawa, The University of Aizu, Computer Science and Engineering, Fukushima 965-8580, Japan

presenting a course to users. The quality of teaching still depends on the actual teacher. To solve this problem comprehensively, an intellectual tutoring system is required.

Online judge systems are environments in which programming can be learned independently [12]. An online judge is an online execution environment that executes and verifies source code submitted by users. Users can improve their programming skills by solving problems in various categories (e.g., syntax, semantics, algorithms, data structures). In addition to automatic verification, an online judge provides a quantitative performance evaluation (execution time, memory usage, etc.) of the program. Although online judges satisfy certain requirements of an autonomous learning environment, it is extremely difficult for novice programmers to continue learning using online judges because they do not provide concrete feedback. Originally, online judges were created for programming contests, in which a solution is either accepted or rejected. Thus, programmers can only recognize that their program has an error, not which type of error (e.g., logic error, compile error, run-time error). For such a program, which is "executable for any input", but "does not satisfy the specifications of the problem", the judgment does not provide useful support for the user. Therefore, novice programmers tend to give up if their logic errors cannot be debugged.

There are various approaches to programming education, including methods that use program execution traces and program dependence graphs. Here, we propose a logic error detection algorithm based on structure patterns, error degree, and abstract syntax trees. A structure pattern is a list generated from the corresponding abstract syntax tree and is an index of similarity at the structure level. An error degree is a value generated from a result of the detection algorithm and is an index of similarity at the characteristic level. The goal of the proposed algorithm is to provide optimal and personalized feedback to individuals. The proposed algorithm provides personalized feedback based on filtering using structure patterns, error degree, and the characteristics of abstract syntax trees.

In this paper, an algorithm for detecting logic errors and the corresponding API, which can be used to construct intelligent coding editors, are presented. The results of quantitative and qualitative experiments are shown. The quantitative experiments are used to verify the detection range of logic errors. The results show that the proposed algorithm can detect logic errors in many cases. The qualitative experiments are used to determine the detection accuracy and evaluate the appropriateness of feedback based on the detection results. The proposed algorithm is shown to detect logic errors and provide appropriate feedback in many cases. In some cases, generated feedback was inappropriate for learning support. This paper is an extension of the work originally presented in "2018 9th International Conference on Awareness Science and Technology (iCAST)" titled "Logic Error Detection Algorithm for Novice Programmers based on Structure Pattern and Error Degree" [13].

The remainder of this paper is organized as follows. Section 2 presents related research. Section 3 describes the logic error detection algorithm. Section 4 describes the LED API. Section 5 presents experiments that use data from an online judge system. Section 6 shows and discusses the experimental results and suggests possible improvements. Finally, Section 7 summarizes the main conclusions.

2 Related Research

The goal of the present study is to construct an autonomous learning environment for novice programmers. However, autonomous learning of programming has not been clearly defined.

Learner autonomy has been mainly considered in the field of foreign language education by researchers such as Holec [14] [15]. Dickinson proposed the concept of autonomous learning, where learners do not rely on teachers, but are themselves responsible for the results, and make all decisions related to learning [16]. Therefore, autonomous learning can be broadly defined as self-learning controlled by the learner.

This definition of autonomous learning assumes that the learning environment is appropriate. However, there is a shortage of teachers in programming education. In the present study, autonomous learning of programming is defined as autonomous learning in an environment without an actual teacher.

Thus far, a number of approaches to support novice programmers have been proposed. In this paper, we present methods that have different targets and structures, and require different educational resources.

Wu et al. designed practice teaching and assessment methods around an online judge system and proposed resolution methods for existing problems in hybrid learning [17]. Semi-automatic evaluation methods that focus on reducing the burden on teachers have also been proposed [18]. In this approach, students are grouped according to the structure of their source code submitted to the online judge, and guidance is provided for each group. Using a network of online judge users, a hybrid method that allows other users to be asked for guidance has been proposed [19]. This approach uses a clustering result based on the similarity of source code. In addition, a method that uses solutions and a language model based on long short-term memory (LSTM) networks for bug detection has been proposed [20].

Fault localization with program execution traces is a well-known technique for detecting logic errors. Using this concept, Su et al. constructed a Moodle-based online learning environment to provide feedback for logic errors [21]. In a study related to this technique, the accuracy of logic error detection using only 20% of the source code and the effectiveness of feedback using a lightweight spectrum-based algorithm were verified [22].

There are approaches that target specific logic errors to support novices. These approaches focus primarily on conditional branches and detect common logic errors, such as careless mistakes in conditional expressions, priority violations, and infinite loops [23] [24].

Targeted at advanced novice programmers, FrenchPress [25] is an eclipse plugin that focuses on programming style and supports programmers who have not yet assimilated the object-oriented paradigm.

In the present study, logic errors are detected by comparing a source code that may include logic errors with correct source codes stored in an online judge system. A comparison of source code is generally performed to detect code clones (i.e., similar or identical fragments of code) [26] [27] [28]. The following four types of method are commonly used to detect code clones.

Text-based methods directly compare source code as text. This type of comparison includes a lot of unique information, such as

variable names. Token-based methods compare tokens obtained by a lexical analysis of source code. Since this type of comparison is not affected by unique information, such as variable names, it has relatively high accuracy. Tree-based methods compare code based on abstract syntax trees obtained by a syntax analyzer. Since parts unrelated to the meaning of the program are omitted from the syntax tree, the structural information of the source code is used for comparison. Semantic-based methods compare code using various information obtained from a parser. A program dependence graph is primarily used. It is possible to detect discontinuous code clones, which are difficult to find using other methods.

The results obtained from text or token comparison are not necessarily consistent with the structure of the source code. Therefore, it is difficult to generate accurate feedback when using text- and token-based methods for detecting logic errors. Since the program dependency graph used in semantic-based approaches has information on data dependency and control dependency, it is possible to perform a comparison based on semantics rather than the description of source code. However, in the isomorphism judgment of graphs, although the semantics of the programs are the same, it is considered difficult to detect logic errors that differ only in internal values. A comparison using abstract syntax trees is based on the structural information of source code and thus can be performed at the syntax level, considering for example *if* statements or *for* statements. Therefore, this method is considered to be suitable for detecting logic errors by a feedback algorithm.

3 Logic Error Detection Algorithm

In the present study, two kinds of source code, namely "wrong code" and "correct code", are used to demonstrate the proposed algorithm. The wrong code is code submitted to an online judge and judged as "Wrong Answer". It is the target for detecting logic errors by the proposed algorithm. The correct code is code judged as "Accepted" and stored in the online judge system. It is the comparison target for detecting logic errors by the proposed algorithm.

3.1 Logic Error

A logic error is a bug in a program that triggers erroneous behavior but does not cause abnormal termination. In other words, logic errors induce behavior not intended by the programmer. Programs that contain logic errors are valid programs that can be compiled and executed. To debug a logic error, the user must determine the cause of the error by tracing variable information during execution using a number of test cases. Table 1 shows some common examples of logic error.

3.2 General Approach of Proposed Algorithm

In the present study, we propose a logic error detection algorithm based on a comparison between wrong and correct code. The proposed algorithm uses a comparison method based on an abstract syntax tree, which is a data structure in which the grammatical structure of source code is represented by finite labeled oriented trees. An abstract syntax tree is recursively defined by a minimum unit, called a syntax, generated by a parser. For example, *if* and *for* statements are recursively defined by a number of constructs.

Table 1: Common examples of logic error

Type of logic error	Cause
Loop condition	Incorrect number of iterations of *for* statement
Conditional branch	Incorrect conditional expression in *if* statement
Output format	Incorrect rounding
Data type	Incorrect data type
Calculation	Incorrect operator in binary expression
Indexing	Wrong index accessed

In the present study, a construct that is recursively defined by other constructs is denoted as a nonterminal construct, and a construct that is not recursively defined by other constructs is denoted as a terminal construct. Examples of nonterminal constructs are *if* statements and conditional expressions, and examples of terminal constructs are character strings and numeric values. Specifically, the algorithm compares abstract syntax trees of the target source code by applying the following steps recursively.

1. If the information on terminal constructs of the same kind is not different, it is not judged as a logic error.
2. If the information on terminal constructs of the same kind is different, it is judged as a logic error.
3. If the constructs are the same kind of nonterminal construct, they are compared to each other.
4. In other cases, a logic error is judged to have occurred.

The algorithm detects constructs that are logic errors. Algorithm 1 shows the pseudocode of the logic error detection algorithm.

The proposed algorithm can compare compilable source code. However, because of the characteristics of abstract syntax trees, a comparison of source code with different tree shapes generates detection results that are unsuitable for detecting logic errors. In addition, even if the shapes of the trees are the same, a comparison of code with different node meanings will generate incorrect detection results. Therefore, the proposed algorithm has two phases.

In the first phase, an appropriate comparison target is selected. The proposed algorithm compares only two source codes and detects logic errors. However, the user may attempt to solve a problem using various approaches. For example, a loop can be implemented using a *for* statement or a *while* statement. If the detection algorithm is applied to the source code of programs that have quite different structures, incorrect results will be obtained. Figures 1 and 2 show examples of a comparison of source code that uses different algorithms. A suitable comparison target is required to detect logic errors.

Algorithm 1 Logic Error Detection

Require:
x = expression in wrong code, x_n = n-th expression of x
y = expression in correct code, y_n = n-th expression of y
Ensure:
E = an array of Logic Error : Array

```
1:  function DETECTION(x, y)
2:      E ⇐ empty array
3:      if x and y are terminal constructs of the same kind then
4:          if x and y are the same information then
5:              return E                                          ▷ Step 1
6:          else
7:              return E ⇐ x                                      ▷ Step 2
8:          end if
9:      else if x and y are nonterminal constructs of the same kind then
10:         if the number of expressions of x and y are the same then
11:             for i = 1 to the number of expressions in x do
12:                 Connect E and DETECTION(x_i, y_i) together    ▷ Step 3
13:             end for
14:             return E
15:         end if
16:     end if
17:     return E ⇐ x                                              ▷ Step 4
18: end function
```

Figure 1: Example of detection with a *for* statement approach

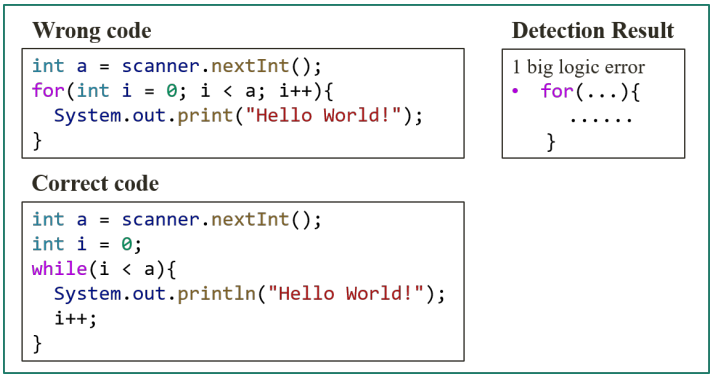


Figure 2: Example of detection with a *while* statement approach

In the second phase, the best detection result is selected. The proposed algorithm generates feedback for the user based on a comparison of the source code. However, the algorithm that the user employs to solve a problem is independent of the source code structure. For example, how conditional branches and libraries are used depends on the user's characteristics. If the detection algorithm is applied to the source code of programs with different algorithms, we may provide feedback while ignoring the characteristics of the corresponding user.

Figures 1 and 3 show examples of a comparison of source code that uses different loop constructions. Therefore, we need to give feedback based on correct codes that has the same characteristics as those of the target codes.

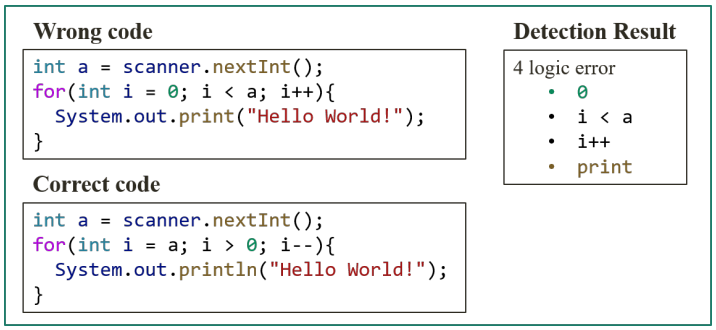


Figure 3: Example of detection with a decrement algorithm

3.3 Structure Pattern

In the present study, structure patterns are used in the first phase. A structure pattern is a list generated from the corresponding abstract syntax tree and is an index of similarity at the structure level. Moreover, it is core data obtained by a depth-first search of the abstract syntax tree and includes a list of expressions and block statements. The generated list contains expressions such as a variable declaration and assignment, and control structures such as *for* and *if* statements. Therefore, a structure pattern can be considered to express the general form of source code.

Figure 4 shows a sample source code. Figure 5 shows an example of an abstract syntax tree parsed from the sample source code. Table 2 shows an
example of a structure pattern generated from the sample source code.

```
int a = sc.nextInt();
int b = sc.nextInt();
if(a < b) {
  System.out.println("a < b");
}else if(a > b) {
  System.out.println("a > b");
}else {
  System.out.println("a == b");
}
```

Figure 4: Sample source code

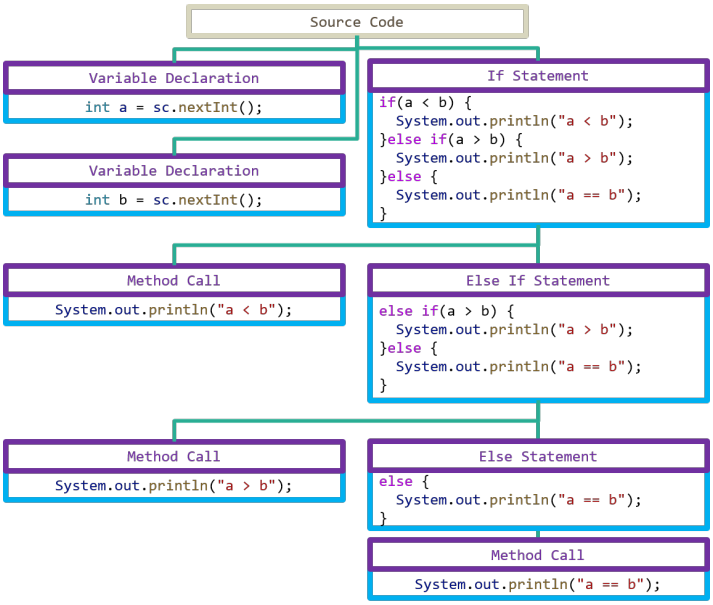


Figure 5: Abstract syntax tree for the source code in Figure 4

Programs with the same structure patterns have the same order and kind of constructs. Therefore, we can obtain appropriate comparison targets by filtering correct code using the structure pattern. For example, the structure patterns for the source code in Figures 1 and 2 are shown in Tables 3 and 4, respectively. As shown, appropriate correct codes for comparison can be obtained by filtering according to the structure pattern.

Table 2: Structure pattern for the source code in Figure 4

Index No.	Syntax
1	Variable Declaration
2	Variable Declaration
3	*If* Statement
4	Method Call
5	*Else* Statement
6	*If* Statement
7	Method Call
8	*Else* Statement
9	Method Call

Table 3: Structure pattern of wrong code and correct code shown in Figure 1

Index No.	Syntax
1	Variable Declaration
2	*For* Statement
3	Method Call

Table 4: Structure pattern of correct code shown in Figure 2

Index No.	Syntax
1	Variable Declaration
2	Variable Declaration
3	*While* Statement
4	Method Call
5	Binary Expression

3.4 Error Degree

In the present study, the error degree is used in the second phase. An error degree is a value generated from a result of the detection algorithm and is an index of similarity at the characteristic level. Moreover, it is a total value obtained by assigning a weight according to the type of detected logic error and can be expressed as (1).

x = the number of logic errors
w = weight of logic error
n = the number of logic error types

$$E_{rrordegree} = \sum_{i=1}^{n} x_i w_i \tag{1}$$

The error degree is zero when the wrong code is compared with itself. An increase in error degree indicates that the source code has characteristics different from those of the wrong code. Therefore, we can obtain optimal feedback by comparing the wrong code with all versions of correct code and by filtering the detection results using the error degree. Table 5 shows the weighting rules. The weight depends on the logic error type in the source code. For example, errors at the element level, such as terminal constructs

and operators, have lower weight, whereas errors at the syntax level, such as the type of syntax and its presence in the code, have higher weight. Errors related to the number of constructs are weighted in proportion to this number.

Table 5: Weighting rules for error degree

Error type	Weight
Terminal construct	1
Operators in expressions	3
Number of constructs	2*constructs
Existence of component	5
Construct that exists only on one side	10
Types of construct for comparison	10

The error degree compensates for the weaknesses of filtering by structure pattern. For example, Figures 6 and 7 show the results of logic error detection for the source code of programs with the same structure patterns presented in Figure 1 and Figure 3. As shown in these figures, more accurate feedback can be obtained by calculating the error degree from the detection result.

Detection Result

Logic Error	Error Type	Weight
print	Terminal construct	1

Error Degree = 1

Figure 6: Detailed detection results for source code in Figure 1

Detection Result

Logic Error	Error Type	Weight
0	Types of construct for comparison	10
i < a	Operators in expressions	3
i ++	Operators in expressions	3
print	Terminal construct	1

Error Degree = 17

Figure 7: Detailed detection result for source code in Figure 3

3.5 Extended Detection Algorithm

The purpose of the proposed algorithm is to support autonomous learning by detecting logic errors in a wrong code and providing the corresponding feedback. The detection algorithm provides feedback through the following steps.

1. Analyze the judged source code and generate a structure pattern.
2. From the data stored in the online judge system, extract correct codes with the same structure pattern.
3. Compare all of the extracted correct codes with the wrong code, and calculate the error degree based on the detection results.
4. Provide information obtained based on the detection result with the lowest error degree as feedback for the user.

Algorithm 2 shows the pseudocode of the extended logic error detection algorithm.

4 Implementation of Algorithm API

Section 3 described the proposed logic error detection algorithm for general online judge systems. In the present study, we also developed an LED web API for the proposed algorithm that is compatible with Aizu Online Judge (AOJ) [29].

4.1 Aizu Online Judge

AOJ is an online judge system developed and operated by the University of Aizu. An online judge system is a service that compiles and executes programs submitted by users and automatically evaluates their accuracy and performance using various test cases and verification machines. At present, AOJ contains source code from 60,000 users and 4 million judgments. AOJ contains various problems, such as those used in programming contests (e.g., ACM-ICPC) and those specific to a certain topic, including Introduction to Programming (ITP), “Algorithm and Data Structures”, and related libraries. For these problems, limit values are set for CPU time and memory usage, and the user can practice by solving the problem under these limitations.

The user obtains a verdict by submitting the answer code for each question to the judge system. The judge system instantly provides feedback (accepted/rejected) based on various test cases for the problem, together with the CPU time and memory usage. The type of failure (Time Limit Exceeded, Memory Limit Exceeded, Runtime Error, etc.) is also given in the feedback. Users can debug their code based on the feedback and resubmit code as many times as desired. In the present study, the aim is to detect logic errors, so we only cover programs corresponding to "Accepted" and "Wrong Answer".

AOJ has an API for accessing a part of its database. The API can be used for developing various tools. For example, there are submitted source code, submission history including log of trial and error.

4.2 Logic Error Detector API

In the present study, we developed an LED that uses source codes stored in AOJ. Furthermore, by implementing peripheral functions, we developed as a web API for the LED.

Algorithm 2 Extended Logic Error Detection

Require:
X = submitted code
Y = array of accepted programs, Y_n = n-th accepted program
Ensure:
F = Feedback
function EXTENDED DETECTION(X) ▷ Step 1
  R = empty array
  $p_{attern} \Leftarrow$ structure pattern of X ▷ Step 2
  $Y \Leftarrow$ all the correct codes submitted for a given problem as X
  $Y \Leftarrow Y$ that have the same structure pattern as p_{attern} ▷ Step 3
  for i = 1 to number of Y **do**
    add detection result based on ALGORITHM 1(X, Y_n) to R ▷ Step 4
  end for
  return $F \Leftarrow$ detection result with the lowest error degree in R ▷ Step 5
end function

From the viewpoint of data availability and practicality, source codes in Java language are analyzed and utilized for the LED API. There are three main reasons for selecting Java. Firstly, it is often adopted as an introduction to object-oriented programming languages for novice programmers. Secondly, it is the second most frequently used language in AOJ. The use of stored data makes it possible to detect logic errors more accurately, so there must be a sufficient number of comparison objects. Thirdly, an open-source library called JavaParser can be used to convert source code written in Java into an abstract syntax tree.

The LED API is implemented using docker containers that serve as a web server, an application server, and a database server. Figure 8 shows an architecture diagram of the LED API.

The web container, based on Nginx, is responsible for front-end processing and acts as a reverse proxy. At the front-end, we implemented tools for research and simple editors using Vue.js and trial of API is possible. The reverse proxy controls access to the application container. Only the web container is exposed to the outside world via the API to prevent attacks on the application and database containers.

The application container is responsible for the back-end processing by the LED, which is the application that implements the proposed algorithm. Since only the host can connect to the database container, it can be used together with the reverse proxy of the web container to prevent attacks on the database container.

LED is a Java application developed using Spring Boot based on Spring Framework. MySQL and Hibernate, as an object-relational mapper, are adopted. JavaParser is used for parsing source code written in Java.

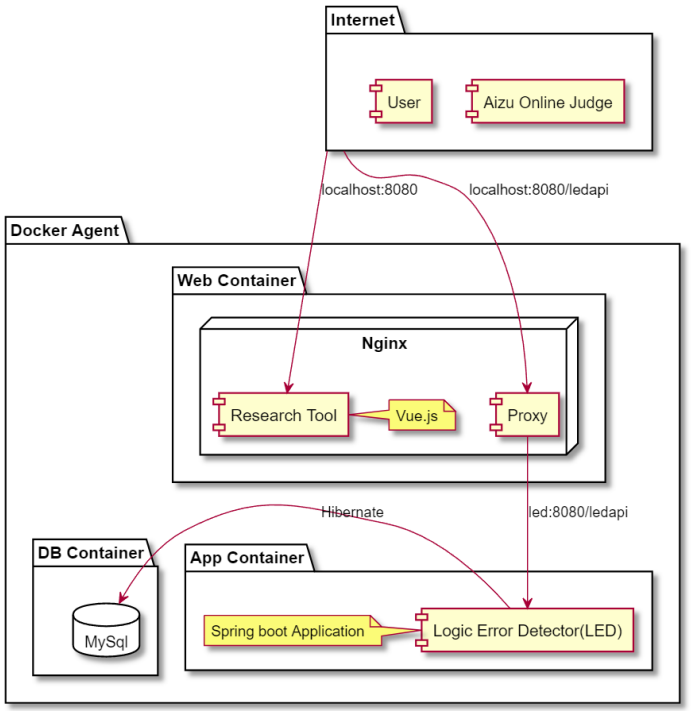


Figure 8: LED architecture diagram

4.3 Class Composition

Figure 9 shows a class diagram of LED. The SourceCode class is used to represent source codes submitted to and stored in AOJ. It has a many-one relation with the Problem, Status, and StructurePattern classes. It includes information on the ID of the user who submitted the code, the ID of the judge, and the submitted code.

The StructurePattern class represents the structure pattern of source code for each program. It has an one-to-many relationship with the SourceCode and PatternComponent classes.

The PatternComponent class is used to represent the type and order of syntax in the structure pattern. It has many-to-one and composition relationships with the StructurePattern class. This is because the structure pattern includes the type and order of syntax, and there are cases where the source code of multiple programs has the same structure patterns, but the pattern component completely depends on the structure pattern.

The ErrorReport class is used to represent the comparison result

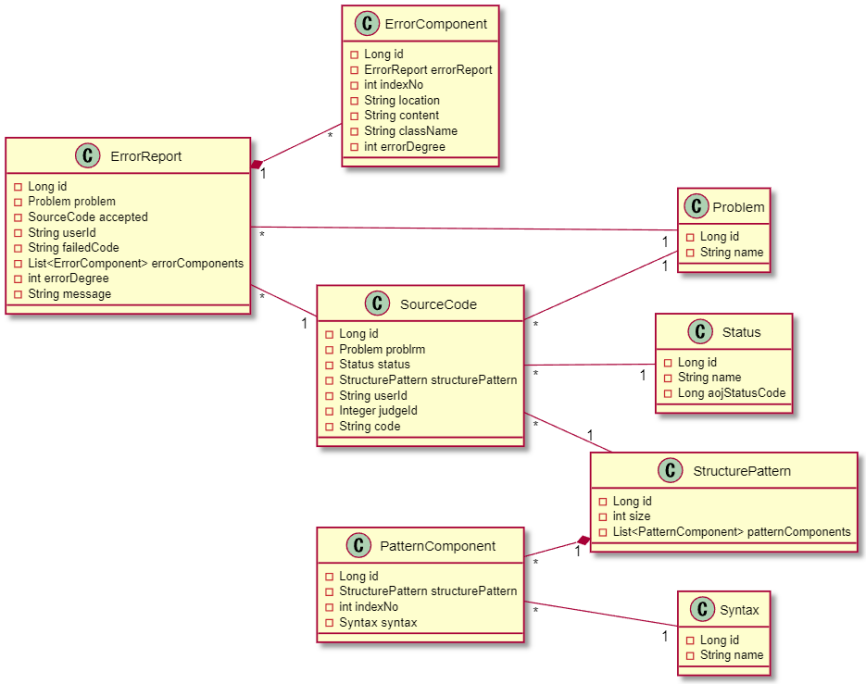


Figure 9: LED class diagram

of the logic error detection algorithm and is also a feedback class of the LED API. It has a many-to-one relationship with the SourceCode and Problem classes. It has many-to-one and composition relationships with the ErrorComponent class. It includes information on wrong codes in which logic errors were detected and correct codes used to generate the detection result selected by filtering using the error degree.

The ErrorComponent class is used to represent details of the detected logic error. It has the location, content, class of syntax, and error degree information of the logic error detected in the wrong code.

4.4 Behavior

Figure 10 shows the relationship between classes and objects. The flow of logic error detection using the LED API is as follows.

1. The user submits a source code that is judged as "Wrong Answer".
2. The wrong code is analyzed and its structure pattern is generated.
3. The structure pattern of the wrong code is searched for in the database.
4. Correct codes with the same structure pattern as that of the wrong code are searched for in the database and used as candidate codes for detection.
5. Using LED, all candidate codes are respectively compared to the wrong code and an error report is generated.
6. The error report with the lowest error degree is provided to the user as feedback.

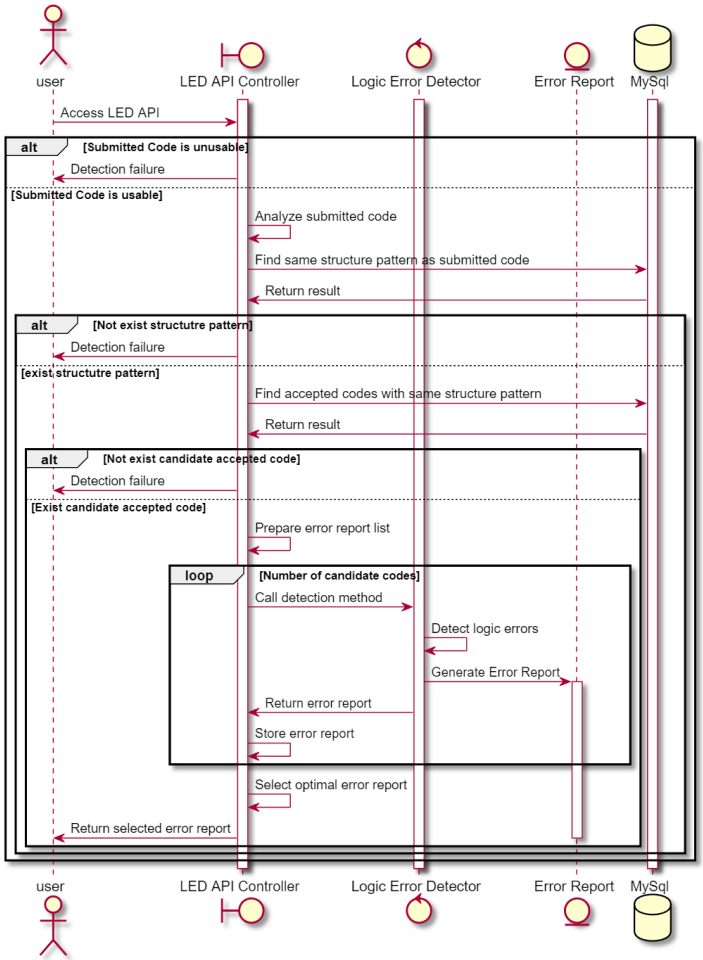


Figure 10: LED sequence diagram

5 Experiment

We conducted quantitative and qualitative experiments using the LED API and AOJ. The LED API is first used to detect logic errors in wrong code submitted by a user who has never solved the problem. Therefore, in this experiment, source codes judged to be "Wrong answer" in AOJ were used as wrong codes.

5.1 Basic Dataset from AOJ

AOJ has a problem set for novices called ITP. In this experiment, all problems in ITP were used. The dataset consists of source code that satisfies the following conditions.

- Submitted to solve the target problems.
- Judged as “Accepted” or “Wrong Answer”.
- Written in Java.
- Consists only of the main method.
- Considered not to be created by the same user.

Tables 6 and 7 show the details of the dataset. “Problem Id” is the problem id of the experimental target. “Usable Programs” is the number of available programs. “Structure Pattern” is the number of structure patterns. “Common Pattern Programs” is the number of programs with the most common structure pattern.

5.2 Experiment 1: Quantitative Method

This subsection describes experiments conducted using quantitative methods.

AOJ stores all the source codes submitted by users. Therefore, there are correct and incorrect answers to certain problems submitted by a given user. By comparing correct codes and the wrong code submitted by a given user, it is possible to obtain reliable feedback *F* necessary for debugging. From the inclusion relation of feedback *F* and feedback *F'* obtained using the LED API, we measure what portion of logic errors can be detected by the proposed algorithm.

In this experiment, we evaluated whether the following statements are true:

- The LED API can be used to detect all logic errors.
- The LED API can be used to detect some logic errors.

For evaluating the first statement, we measured whether *F'* completely includes *F*. The code debugged based on information provided by feedback *F* has already been accepted by AOJ. Therefore, when feedback *F'* completely includes feedback *F*, it can be regarded that all the logic errors were detected.

For evaluating the second statement, we measured whether *F'* contains some of *F*. During debugging, a user may modify not only the algorithm, but also variable names, the order of processing, etc. Therefore, when some of feedback *F* are included in feedback *F'*, it can be regarded that some logic errors were detected.

The following procedures were applied to each target problem.

1. Apply the proposed algorithm to correct code and wrong code submitted by a given user and obtain feedback *F*.
2. Generate feedback *F'* from the wrong code using the LED API.
3. Determine the inclusion relation of *F* and *F'*.

The dataset consists of source code that satisfies the following conditions.

- Correct code and wrong code that submitted by the same user.
- These two programs have the same structure patterns.

5.3 Experiment 2: Qualitative method

This subsection describes experiments conducted using qualitative methods.

In Experiment 1, the accuracy of logic error detection was investigated using a quantitative method. However, the experiment only measured the detection range, and it is not known whether the feedback is sufficient for debugging. Therefore, we manually debugged a code using feedback from the LED API and verified whether the wrong code was acceptable by AOJ. In this experiment, we evaluated whether the following statements are true:

- The LED API can be used to correctly detect logic errors.
- The provided feedback is appropriate for supporting learning.

The following procedures were applied to each target problem.

1. Randomly select a wrong code in AOJ.
2. Detect logic errors using the LED API.
3. Debug the wrong code manually using feedback and verify whether the debugged code is acceptable by AOJ.

Table 6: Problem Information

Problem ID	Topic	Summary of Problem
ITP1_1_A	Getting Started	Print "Hello World" to standard output.
ITP1_1_B		Calculate the cube of a given integer x
ITP1_1_C		Calculate the area and perimeter of a given rectangle.
ITP1_1_D		Read an integer S [second] and convert it to h:m:s.
ITP1_2_A	Branch on Condition	Print small/large/equal relation of given two integers a and b.
ITP1_2_B		Read three integers a, b and c, and print "Yes" if $a < b < c$, otherwise "No".
ITP1_2_C		Read integers and print them in ascending order.
ITP1_2_D		Read a rectangle and a circle, and determine whether the circle is arranged inside the rectangle.
ITP1_3_A	Repetitive Processing	Print "Hello World" 1000 times.
ITP1_3_B		Read an integer x and print it as is for multiple test cases.
ITP1_3_C		Read two integers x and y, and print them in ascending order for multiple test cases.
ITP1_3_D		Read three integers a, b, and c, and print the number of divisors of c in [a, b].
ITP1_4_A	Computation	Read two integers a and b, and calculate a / b in different data types.
ITP1_4_B		Calculate the area and circumference of a circle for given radius r.
ITP1_4_C		Read two integers, a and b, and an operator op, and then print the value of a op b.
ITP1_4_D		Read a sequence of n integers $a_i(i=1,2,...n)$, and print the minimum value, maximum value, and sum of the sequence.
ITP1_5_A	Structured Program I	Draw a rectangle whose height is H cm and width is W cm.
ITP1_5_B		Draw a frame whose height is H cm and width is W cm.
ITP1_5_C		Draw a chessboard whose height is H cm and width is W cm.
ITP1_5_D		Structured programming without *goto* statement.
ITP1_6_A	Array	Read a sequence and print it in reverse order.
ITP1_6_B		For given a deck of cards, find missing cards.
ITP1_6_C		Count the number of elements in a three-dimensional array.
ITP1_6_D		Read a n×m matrix A and a m×1 vector b, and print their product Ab.
ITP1_7_A	Structured Program II	Read a list of student scores and evaluate the grade for each student.
ITP1_7_B		Find the number of combinations of integers that satisfy 1) three distinct integers from 1 to n, and 2) sum of the integers is x.
ITP1_7_C		Read the number of rows r, columns c, and a table of r×c elements, and print a new table that includes the total sum for each row and column.
ITP1_7_D		Read a n×m matrix A and a m×l matrix B, and print their product.
ITP1_8_A	Character	Convert uppercase/lowercase letters to lowercase/uppercase.
ITP1_8_B		Read an integer and print the sum of its digits.
ITP1_8_C		Count the number of appearances for each letter (ignoring case).
ITP1_8_D		Find a pattern p in ring-shaped text s.
ITP1_9_A	String	Read a word W and text T, and print the number of times that word W appears in text T.
ITP1_9_B		Read a deck (a string) and shuffle it, and print the final state (a string).
ITP1_9_C		Read a sequence of Taro and Hanako cards and report the final scores of the game.
ITP1_9_D		Perform a sequence of commands including reverse and replace operations.
ITP1_10_A	Math Functions	Calculate the distance between two points $P1$($x1$, $y1$) and $P2$($x2$, $y2$).
ITP1_10_B		For the given two sides of a triangle a and b and the angle C between them, calculate the area, circumference, and height of the triangle.
ITP1_10_C		Calculate the standard deviation.
ITP1_10_D		Read two n-dimensional vectors x and y, and calculate Minkowski's distance, where p=$1,2,3$, ∞.
ITP1_11_A	Structure and Class	Simulate rolling a dice.
ITP1_11_B		Print the integer on the right side face of the rolled die.
ITP1_11_C		Read two dice and determine whether they are identical.
ITP1_11_D		Read n dice and determine whether they are all different.

6 Results and Discussion

In this section, we present and discuss the results of each experiment.

6.1 Results of Experiment 1

Table 8 shows the results of Experiment 1. "Problem Id" is the problem id of the experimental target. "All Users" is the number of users who submitted source code. "Usable Users" is the number

Table 7: Statistics of the Dataset Used in the Experiment

Problem ID	Number of Programs	Usable Programs	Structure Pattern	Common Pattern Programs
ITP1_1_A	5132	5094	5	5030
ITP1_1_B	4408	3713	50	1274
ITP1_1_C	3292	2743	56	744
ITP1_1_D	2365	1832	68	504
ITP1_2_A	2836	2231	82	629
ITP1_2_B	2283	1683	68	619
ITP1_2_C	2779	1099	71	234
ITP1_2_D	2919	1682	60	628
ITP1_3_A	2290	2194	15	1436
ITP1_3_B	3180	1303	90	71
ITP1_3_C	2538	727	75	67
ITP1_3_D	1622	983	58	261
ITP1_4_A	3715	2605	66	541
ITP1_4_B	2263	1825	56	316
ITP1_4_C	1506	347	49	35
ITP1_4_D	3027	735	48	54
ITP1_5_A	1477	701	65	107
ITP1_5_B	1404	354	46	57
ITP1_5_C	1236	421	53	58
ITP1_5_D	3120	717	58	147
ITP1_6_A	1345	447	46	53
ITP1_6_B	1027	181	25	23
ITP1_6_C	1340	212	19	55
ITP1_6_D	743	295	32	36
ITP1_7_A	1683	196	29	28
ITP1_7_B	1287	230	28	33
ITP1_7_C	839	109	19	14
ITP1_7_D	1207	216	24	37
ITP1_8_A	927	329	44	23
ITP1_8_B	797	227	30	32
ITP1_8_C	1444	152	23	19
ITP1_8_D	721	233	23	44
ITP1_9_A	1409	270	27	35
ITP1_9_B	642	151	17	28
ITP1_9_C	710	174	23	29
ITP1_9_D	613	74	11	17
ITP1_10_A	594	387	28	98
ITP1_10_B	670	339	24	84
ITP1_10_C	935	78	11	17
ITP1_10_D	420	61	8	14
ITP1_11_A	135	20	4	11
ITP1_11_B	109	18	4	9
ITP1_11_C	108	11	2	8
ITP1_11_D	19	0	0	0

of users who submitted both correct and wrong code with the same structure patterns (and thus logic errors can be detected using the LED API). "Whole Success Users" is the number of users for whom all logic errors were detected using the LED API. In other words, this is the result of the first statement. "Partial Success Users" is the number of users for whom some logic errors were detected using the LED API. In other words, this is the result of the second statement. "Whole Success Ratio" is the ratio of Whole Success Users to Usable Users. "Partial Success Ratio" is the ratio of Partial Success Users to Usable Users. "Difference Success Ratio" is the difference between Whole Success Rate and Partial Success Rate.

The first item is the complete detection rate of logic errors. The results of the experiment show that all logic errors were detected for many problems. Overall, we succeeded in detecting logic errors in more than 70% of programs.

The second item is the partial detection rate of logic errors. The results of the experiments show that a higher detection ratio for partial logic errors was achieved for many problems. Overall, we succeeded in detecting logic errors in more than 80% of programs.

6.2 Results of Experiment 2

Table 9 shows the results of Experiment 2. The first item is the detection accuracy of logic errors. Considering the features of the evaluation method, it was considered that logic errors could not be correctly detected when the source code could be modified only by factors outside the feedback range. The results of the experiment show that we successfully detected logic errors in 100% of test cases. Therefore, the proposed detection algorithm has sufficient ability to detect logic errors in wrong code.

The second item is related to the appropriateness of feedback. The experimental results show that we succeeded in providing appropriate feedback in 80.4% of the test cases. This evaluation item is based on the qualitative method used to verify the appropriateness for supporting learning. Therefore, we should discuss the appropriateness of the judgment.

6.3 Discussion

In this subsection, we discuss the experimental results. Some ideas for improvement based on the experimental results are also given.

6.3.1 Discussion for Experiment 1

The results of Experiment 1 confirm that the proposed detection algorithm can support users who attempt to solve problems. We also confirmed that the detection algorithm can detect all the logic errors in a wrong code in many cases. However, results for statement 2 indicate that feedback may contain syntax that is unrelated to the algorithm. Therefore, it is necessary to improve detection accuracy by improving the detection algorithm.

6.3.2 Discussion for Experiment 2

The results of Experiment 2 confirm that logic errors can be detected by the proposed detection algorithm. In addition, we confirmed that the detection algorithm can provide appropriate feedback in many cases. Therefore, it is necessary to generalize some specific cases and to improve the algorithm.

6.3.3 Discussion for All Experiments

The results of the two experiments confirmed that the proposed algorithm can accurately detect logic errors and provide appropriate feedback. The proposed algorithm detects logic errors using source codes stored in the online judge system. To increase accuracy, we filter the codes using structure patterns and error degree. Therefore, detection accuracy depends on the diversity of source code stored in the online judge system. In the present study, although we only considered codes written in Java, the proposed algorithm can be applied to all context-free languages. There is more source code written in languages such as C and C ++, and thus logic errors can be detected with higher accuracy.

6.3.4 Improvements

In this subsubsection, we suggest some improvements based on the experimental results.

Based on the results of Experiment 1, we suggest the following two improvements. One improvement is the removal of elements not related to logic errors detected by the detection algorithm. For example, elements with the same meaning in different formats, such as variable names and logical expressions, can be removed because they are not related to logic errors. Another improvement is the use of a comparison method that is independent of the structure pattern (i.e., generalization of the algorithm). In Experiment 1, a rule regarding the structure pattern as a condition of the dataset was used. This rule was used because the LED API does not have a function for comparing
source codes with different structure patterns. Therefore, we suggest generalizing the algorithm by unifying the variable declaration part and determining the format of the calculation formula. This generalization would make it possible to compare source codes with different structure patterns without changing the meaning of the algorithm.

Based on the results of Experiment 2, we suggest the following three improvements. The first improvement is a comparison method of conditional branching. The conditional branching algorithm strongly depends on the logical expression that is employed by the user. In particular, filtering by structure pattern is not effective for problems that require multiple standard outputs using conditional branching. Therefore, we suggest a comparison based on a logical expression and feedback regarding the processing order. The second improvement is a comparison method of binary expressions. Because binary expressions are parsed using certain rules, a comparison between binary expressions from different structures are not effective. Therefore, we suggest making the error degree more accurate by implementing special logic to compare binary expressions. The third improvement is a review of the weighting when the type of syntax is the same but the elements are different. As an example, method call expressions can be very different, even using the same kind of syntax.

Table 8: Results of Experiment 1

	Users		Success Users		Success Rate		
Problem ID	All	Usable	Whole	Partial	Whole	Partial	Difference
ITP1_1_A	3502	358	335	358	93.58%	100.00%	+6.42%
ITP1_1_B	2425	144	110	121	76.39%	84.03%	+7.64%
ITP1_1_C	2006	187	114	142	60.96%	75.94%	+14.97%
ITP1_1_D	1618	130	86	102	66.15%	78.46%	+12.31%
ITP1_2_A	1692	291	249	263	85.57%	90.38%	+4.81%
ITP1_2_B	1525	126	106	112	84.13%	88.89%	+4.76%
ITP1_2_C	1505	87	49	74	56.32%	85.06%	+28.74%
ITP1_2_D	1194	261	93	111	35.63%	42.53%	+6.90%
ITP1_3_A	1542	270	201	208	74.44%	77.04%	+2.59%
ITP1_3_B	1405	129	108	114	83.72%	88.37%	+4.65%
ITP1_3_C	1199	63	55	57	87.30%	90.48%	+3.17%
ITP1_3_D	1059	105	92	97	87.62%	92.38%	+4.76%
ITP1_4_A	1107	315	180	252	57.14%	80.00%	+22.86%
ITP1_4_B	1134	250	157	184	62.80%	73.60%	+10.80%
ITP1_4_C	1022	14	14	14	100.00%	100.00%	0%
ITP1_4_D	909	176	107	147	60.80%	83.52%	+22.73%
ITP1_5_A	977	24	21	22	87.50%	91.67%	+4.17%
ITP1_5_B	911	33	22	26	66.67%	78.79%	+12.12%
ITP1_5_C	880	20	14	16	70.00%	80.00%	+10.00%
ITP1_5_D	968	58	51	53	87.93%	91.38%	+3.45%
ITP1_6_A	831	17	15	16	88.24%	94.12%	+5.88%
ITP1_6_B	656	15	12	14	80.00%	93.33%	+13.33%
ITP1_6_C	623	33	26	27	78.79%	81.82%	+3.03%
ITP1_6_D	566	11	9	11	81.82%	100.00%	+18.18%
ITP1_7_A	719	21	20	20	95.24%	95.24%	0%
ITP1_7_B	643	24	21	23	87.50%	95.83%	+8.33%
ITP1_7_C	595	3	3	3	100.00%	100.00%	0%
ITP1_7_D	484	66	49	64	74.24%	96.97%	+22.73%
ITP1_8_A	636	14	13	14	92.86%	100.00%	+7.14%
ITP1_8_B	591	2	0	1	0.00%	50.00%	+50.00%
ITP1_8_C	525	5	5	5	100.00%	100.00%	0%
ITP1_8_D	448	11	7	8	63.64%	72.73%	+9.09%
ITP1_9_A	518	52	35	42	67.31%	80.77%	+13.46%
ITP1_9_B	456	7	7	7	100.00%	100.00%	0%
ITP1_9_C	441	6	2	5	33.33%	83.33%	+50.00%
ITP1_9_D	323	4	4	4	100.00%	100.00%	0%
ITP1_10_A	432	32	15	22	46.88%	68.75%	+21.88%
ITP1_10_B	392	38	20	31	52.63%	81.58%	+28.95%
ITP1_10_C	458	14	9	10	64.29%	71.43%	+7.14%
ITP1_10_D	279	2	2	2	100.00%	100.00%	0%
ITP1_11_A	84	1	1	1	100.00%	100.00%	0%
ITP1_11_B	64	1	1	1	100.00%	100.00%	0%
ITP1_11_C	32	1	1	1	100.00%	100.00%	0%
ITP1_11_D	13	0	0	0	0.00%	0.00%	0%

Table 9: Results of Experiment 2

Feedback Type	Reason for Judgment	Cause of Logic Error	Rate	Total
Appropriate	Can be debugged	Literal	6.5%	80.43%
		Output format	19.6%	
		Binary expression	15.2%	
		Loop condition	8.7%	
		Type and output format	2.2%	
		Conditional branch	19.6%	
		Used type	4.3%	
		Used method	4.3%	
Inappropriate	Direct answer	Position of variables	17.4%	19.57%
	Ignored the algorithm	Non-optimal feedback	2.2%	
Wrong	Cannot be debugged		0.0%	0.0%

However, even with such a comparison, the error degree is a very small numerical value. In particular, standard output methods and various library methods are sometimes compared, so that optimal feedback cannot be generated in some cases. Therefore, we suggest measuring to what extent the syntax has the same meaning by using the namespace and arguments. We also suggest detecting the parent syntax itself as a logic error, such as when only variable names are detected as logic errors.

7 Conclusion

This study proposed a logic error detection algorithm based on structure patterns and error degree to support novice programmers. The proposed algorithm uses data in an online judge system to detect logic errors in wrong code and provides feedback for supporting problem solving. In addition, we developed the LED web API based on the proposed algorithm. The LED API is compatible with the online judge system, so it can promote autonomous learning with the corresponding environment.

Two experiments were conducted to verified the accuracy of the proposed algorithm. The experimental results show that the proposed algorithm can accurately detect logic errors and provide appropriate feedback. The proposed algorithm detected logic errors in programs for all the problems in the Introduction to Programming 1 set with high accuracy. In addition, the detection accuracy depends on the amount of data in the online judge system.

Although we constructed the autonomous learning environment and detection algorithm in Java, the proposed algorithm can be applied to various programming languages, such as C/C++, Python, and Ruby. Although the presented LED API is oriented to detec logic errors, the algorithm is capable of suggesting possible solutions. So, our future works include the extension of the API after conducting further experiments.

References

[1] C. B. Frey and M. A. Osborne, "The future of employment: How susceptible are jobs to computerisation?" 2013. [Online]. Available: https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf

[2] I. Koichi and T. Yuta, "Digitization, Computerization, Networking, Automation, and the Future of Jobs in Japan (article in Japanese with an abstract in English)," Research Institute of Economy, Trade and Industry (RIETI), Policy Discussion Papers (Japanese) 18009, May 2018. [Online]. Available: https://ideas.repec.org/p/eti/rpdpjp/18009.html

[3] W. E. Forum, "The future of jobs report 2018," 2018. [Online]. Available: http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf

[4] J. Gal-Ezer, C. Beeri, D. Harel, and A. Yehudai, "A high school program in computer science," *Computer*, vol. 28, no. 10, pp. 73–80, Oct. 1995. [Online]. Available: http://dx.doi.org/10.1109/2.467599

[5] S. Furber *et al.*, "Shut down or restart? the way forward for computing in uk schools," *The Royal Society, London*, 2012. [Online]. Available: https://royalsociety.org/~/media/education/computing-in-schools/2012-01-12-computing-in-schools.pdf

[6] Y. Okajima, "Research on the impact of compulsory programming education," *Chuo University Policy Research Institute Annual Report*, no. 21/2017, pp. 203 – 218, 2018-08-23. [Online]. Available: http://id.nii.ac.jp/0341/00005893

[7] K. Komatsu, "Problems and countermeasures of programming education (article in japanese)," *Business review, Faculty of Business Administration, Bunkyo Gakuin University*, vol. 25, no. 1, pp. 83–104, dec 2015. [Online]. Available: https://ci.nii.ac.jp/naid/40020853834/en/

[8] S. Furber *et al.*, "After the reboot: computing education in uk schools," *The Royal Society, London*, 2017. [Online]. Available: https://royalsociety.org/~/media/policy/projects/computing-education/computing-education-report.pdf

[9] C. M. Intisar and Y. Watanobe, "Classification of online judge programmers based on rule extraction from self organizing feature map," in *2018 9th International Conference on Awareness Science and Technology (iCAST)*, Sep. 2018, pp. 313–318.

[10] T. Saito and Y. Watanobe, "Learning path recommender system based on recurrent neural network," in *2018 9th International Conference on Awareness Science and Technology (iCAST)*, Sep. 2018, pp. 324–329.

[11] C. M. Intisar and Y. Watanobe, "Cluster analysis to estimate the difficulty of programming problems," in *Proceedings of the 3rd International Conference on Applications in Information Technology(ICAIT)*, 2018.

[12] S. Wasik, M. Antczak, J. Badura, A. Laskowski, and T. Sternal, "A survey on online judge systems and their applications," *ACM Comput. Surv.*, vol. 51, no. 1, pp. 3:1–3:34, Jan. 2018. [Online]. Available: http://doi.acm.org/10.1145/3143560

[13] Y. Yoshizawa and Y. Watanobe, "Logic error detection algorithm for novice programmers based on structure pattern and error degree," in *2018 9th International Conference on Awareness Science and Technology (iCAST)*, Sep. 2018, pp. 297–301.

[14] H. Holec, *Autonomy and foreign language learning.* ERIC, 1979.

[15] D. G. Little, *Learner autonomy: Definitions, issues and problems.* Authentik Language Learning Resources, 1991.

[16] L. Dickinson, "Autonomy and motivation a literature review," *System*, vol. 23, no. 2, pp. 165–174, 1995.

[17] H. Wu, Y. Liu, L. Qiu, and Y. Liu, "Online judge system and its applications in c language teaching," in *2016 International Symposium on Educational Technology (ISET)*, July 2016, pp. 57–60.

[18] S. Buyrukoglu, F. Batmaz, and R. Lock, "Increasing the similarity of programming code structures to accelerate the marking process in a new semi-automated assessment approach," in *2016 11th International Conference on Computer Science Education (ICCSE)*, Aug 2016, pp. 371–376.

[19] E. Stankov, M. Jovanov, B. Kostadinov, and A. M. Bogdanova, "A new model for collaborative learning of programming using source code similarity detection," in *2015 IEEE Global Engineering Education Conference (EDUCON)*, March 2015, pp. 709–715.

[20] Y. Teshima and Y. Watanobe, "Bug detection based on lstm networks and solution codes," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct 2018, pp. 3541–3546.

[21] X. Su, J. Qiu, T. Wang, and L. Zhao, "Optimization and improvements of a moodle-based online learning system for c programming," in *2016 IEEE Frontiers in Education Conference (FIE)*, Oct 2016, pp. 1–8.

[22] E. Araujo, M. Gaudencio, D. Serey, and J. Figueiredo, "Applying spectrum-based fault localization on novice's programs," in *2016 IEEE Frontiers in Education Conference (FIE)*, Oct 2016, pp. 1–8.

[23] T. Nguyen and C. Chua, "A logical error detector for novice php programmers," in *2014 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, July 2014, pp. 215–216.

[24] K. K. Sharma, K. Banerjee, I. Vikas, and C. Mandal, "Automated checking of the violation of precedence of conditions in else-if constructs in students' programs," in *2014 IEEE International Conference on MOOC, Innovation and Technology in Education (MITE)*, Dec 2014, pp. 201–204.

[25] H. Blau, S. Kolovson, W. R. Adrion, and R. Moll, "Automated style feedback for advanced beginner java programmers," in *2016 IEEE Frontiers in Education Conference (FIE)*, Oct 2016, pp. 1–9.

[26] J. Feng, B. Cui, and K. Xia, "A code comparison algorithm based on ast for plagiarism detection," in *2013 Fourth International Conference on Emerging Intelligent Data and Web Technologies*, Sept 2013, pp. 393–397.

[27] G. Tao, D. Guowei, Q. Hu, and C. Baojiang, "Improved plagiarism detection algorithm based on abstract syntax tree," in *2013 Fourth International Conference on Emerging Intelligent Data and Web Technologies*, Sept 2013, pp. 714–719.

[28] Y. Zhang, X. Gao, C. Bian, D. Ma, and B. Cui, "Homologous detection based on text, token and abstract syntax tree comparison," in *2010 IEEE International Conference on Information Theory and Information Security*, Dec 2010, pp. 70–75.

[29] Y. Watanobe, "Development and operation of an online judge system," *Information Processing*, vol. 56, no. 10, pp. 998–1005, sep 2015. [Online]. Available: https://ci.nii.ac.jp/naid/40020591038/

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 5, 16-22 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Melanoma detection using color and texture features in computer vision systems

Antonio Fuduli[1], Pierangelo Veltri[2], Eugenio Vocaturo[3], Ester Zumpano[*,3]

[1]*Department of Mathematics and Computer Science, University of Calabria, 87036, Italy*

[2]*DSCM - Bioinformatics Laboratory Surgical and Medical Science Department, University Magna Graecia, 88100, Italy*

[3]*DIMES - Department of Computer Science, Modelling, Electronic and Systems Engineering, University of Calabria, 87036, Italy*

ARTICLE INFO

Article history:
Received: 14 June, 2019
Accepted: 22 August, 2019
Online: 06 September, 2019

Keywords:
Melanoma detection
Feature selection
Classification

ABSTRACT

All forms of skin cancer are becoming widespread. These forms of cancer, and melanoma in particular, are insidious and aggressive and if not treated promptly can be lethal to humans. Effective treatment of skin lesions depends strongly on the timeliness of the diagnosis: for this reason, artificial vision systems are required to play a crucial role in supporting the diagnosis of skin lesions. This work offers insights into the state of the art in the field of melanoma image classification. We include a numerical section where a preliminary analysis of some classification techniques is performed, using color and texture features on a data set constituted by plain photographies, to which no pre-processing technique has been applied. This is motivated by the necessity to open new horizons in creating self-diagnosis systems for accessible skin lesions, due also to a huge innovation of cameras, smartphones technology and wearable devices.

1 Introduction

The World Health Organization recognizes skin cancer as one of the most lethal cancers that is spreading widely in many parts of the world in recent decades. Globally, each year between 2 and 3 million of non-melanoma skin cancers and 132,000 of melanoma skin cancers occur [1]. Cutaneous melanoma is still incurable. A successful treatment of skin lesion is heavily dependent on the timeliness of the diagnosis: it is crucial to detect and remove the skin lesion in the early stage. One of the major allert of this tumor is that, in the early stages, melanoma remains hidden, showing itself to be similar to other benign lesions. This aspect, combined with the aggressiveness and rapidity with which it expands, creates difficulties even for expert dermatologists; in addition, in its early phase, it is often underestimated [2]. Dermatologists, in clinical evaluation, in order to identify the tumor lesions of the skin, adhere to some reference protocols. Generally, they refer to the ABCDE protocol [3], the 7-point checklist [4] and the Menzies method [5]. If the investigations conducted leave doubts about the nature of the lesion, the specialist prescribes an analysis of the affected tissue by biopsy. Even today, the reliability of the diagnosis of melanoma can be improved with the support of latest-generation automated analysis tools. Although the human factor continues to be the main discriminating element for the diagnostic accuracy [2], on the other hand such tools could play a relevant role to support the specialist in the diagnosis procedure.
Therefore if, on one hand, it is evident the need to design sophisticated tools supporting the specialists in providing early and accurate diagnosis, on the other hand it is crucial also the necessity to develop low cost solutions alerting in case of suspicious skin lesions. Technology is driving a huge innovation in the healthcare domain and artificial intelligence is playing a key role in providing tools and applications supporting physicians in diagnosis and patient treatment. In particular, computer vision systems are revealing to be fundamental in medical image analysis, since they allow to have non-invasive diagnoses. The specific frameworks for the automatic analysis of skin lesions are constituted by the following fundamental steps: image acquisition, image processing and analysis, features extraction and classification.

This paper is an extension of work originally presented at the 9th International Conference on Information, Intelligence, Systems and Applications (IISA 2018). The goal of the original paper [6] was to offer insights into the state of the art in the field of melanoma image classification by both providing a review of the various features used in the elaboration of computer vision systems and focusing on those ones which appear to be the mostly used on the basis of the recent literature. Here a preliminary comparison between two different approaches (Support Vector Machine and Multiple Instance Learning)

[*]Corresponding Author:Ester Zumpano, DIMES - Department of Computer Science, Modelling, Electronic and Systems Engineering, University of Calabria, 87036, Italy & zumpano@dimes.unical.it

is proposed, pointing out the key role played by the feature selection (color and texture). In particular, this study was inspired by the good results obtained in [7], where the authors applied Multiple Instance Learning techniques for classifying some medical dermoscopic images drawn from the Ph^2 database [8], by using only color features. Our numerical experimentations, reported in this paper, showed that using only color features is not at all high-performance when the classification process is performed on a data base constituted by common plain photographies. On the other hand using in addition texture features provides reasonable classification results, especially if we consider that no pre-processing techniques were used.

This paper is organized as follows. In Section 2 we present the pathogenic mechanisms of the skin cancer, while in Section 3 we recall the main steps adopted in biomedical image processing. In Section 4, a preliminary numerical study is performed for classifying a set of 200 plain photographies (100 melanomas and 100 common nevi), using color and color/texture features. Finally some conclusions are drawn in Section 5, looking at possible future developments.

2 Skin lesions

The first layer of the skin is the epidermis, which is constituted by different types of cells, each of them characterized by specific biological chemical functions: *keratinocytes* have a mechanical function of protection from the external environment, *basal cells* have a reproductional function of the cells of the most superficial layer and *melanocytes* provide protection against ultraviolet rays throught the production of melanin.

Dermis is the second layer of the skin, and is constituted by dermal papillary and reticular dermis. It is located below the epidermis and has a thicker consistency (Figure 1). The third layer of the skin is the subcutaneous layer or hypodermis, and, as the dermis, contains many blood and lymphatic vessels and protects the underlying tissues and the internal organs [9].

The different biological structure of the various layers of the skin implies their specific optical properties. When a beam of white light hits the skin, one part is reflected, another part penetrates into the superficial layers and the last part is absorbed. The specific amount of light that is reflected or absorbed depends on genetic factors of the population including the specific phototype to which the individual belongs. Exposure to light triggers melanocytes that produce melanin which is activated as a particularly effective protective filter on shades of blue light.

Normally the melanin deposits are placed in the epidermis (nevi); in the case of malignant lesions, such as the feared melanoma, melanocytes produce excess melanin and these deposits tend to invade the underlying levels of the skin. The presence of melanin in the dermis is one of the most indicative factors in the presence of melanoma (Figure 2).

Further factors indicative of a malignant lesion are the increased blood flow at the periphery of the lesion, evidence of epithelial desquation and the low flow of blood within the lesion which appears with a bluish red hue. Precisely the color features provide clinical evidence for a suspicion of skin cancer. At this point the specialist prescribes a histology to make an explicit diagnosis. Figure 3 shows a common nevus (a), a dysplastic nevus (b) (which is a benign lesion) and finally a melanoma (c) [8].

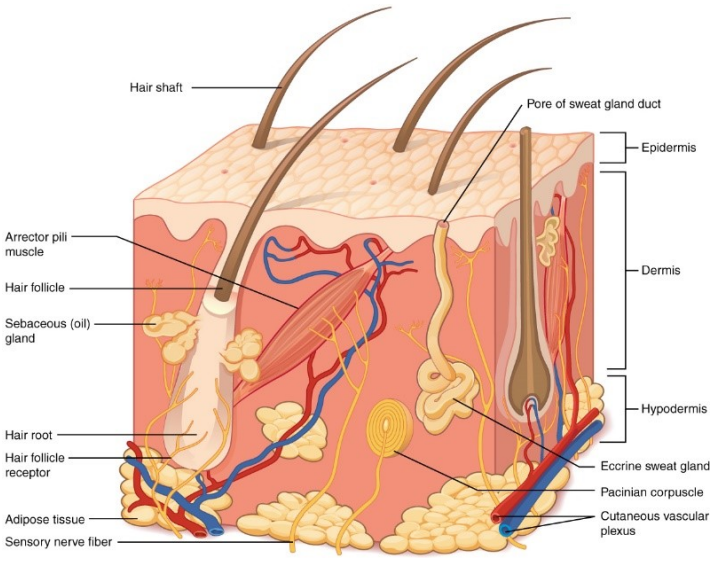


Figure 1: Skin layers

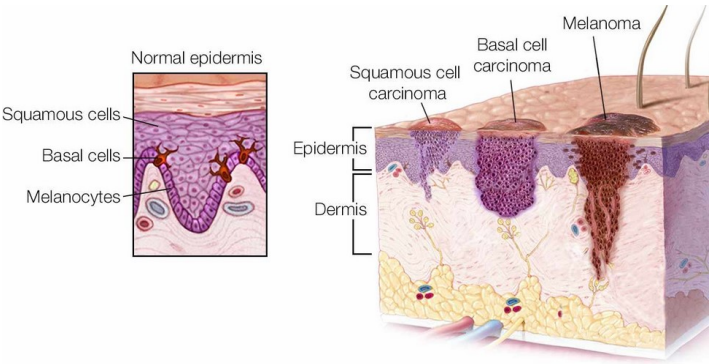


Figure 2: Types of skin cancer

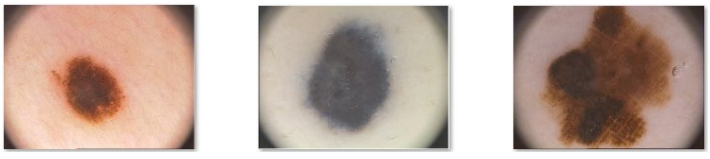


Figure 3: Dermoscopic images of common nevus (a), dysplastic nevus (b) and melanoma (c)

3 Main steps in computer vision systems

In this section we will focus on the fundamental phases that characterize the automatic frameworks for image analysis. We will also refer to the most widespread techniques that characterize the various steps of an artificial vision system: acquisition, pre-processing, segmentation, extraction and selection of features (Figure 4). The classification would be the final step, which is not the main object

of this document, apart from some brief discussion in the numerical section.

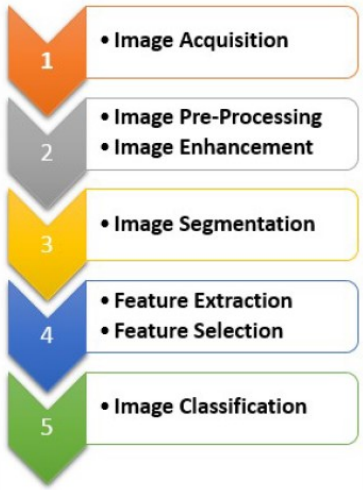


Figure 4: Main steps in biomedical image processing

3.1 Image acquisition

The first step of computer vision systems is the acquisition of the digital image. The proliferation of techniques as well as the development of new software paradigms has found its way into the numerous techniques that have been presented by researchers and industrial stakeholders. Based on the corresponding pros and cons, each techniques results more suitable for specific cases and less for other ones. Among the mainly widespread techniques we report total cutaneous photography, dermoscopy, ultrasound, magnetic resonance imaging (MRI), confocal scanning laser microscopy (CSLM), positron emission tomography (PET), optical coherence tomography (OCT), multispectral imaging, computed tomography (CT), multifrequency electrical impedance and Raman spectra. The epiluminescence microscopy (ELM), also known as dermoscopy, allows an in-depth analysis of the surface texture of skin lesions as the technique makes the epidermis translucent. Transmission electron microscopy (TEM) is generally used to monitor the evolution of melanoma. Through an appropriate light source the layers of the skin become translucent and this makes possible to appreciate the vascularization of the lesion also analyzing the colorimetric variations of the pigmentation. Computed tomography (CT) is also used to assess the patient's subjective response, which also allows diagnosis to be made on suspicious lesions.

Positron emission tomography (PET) is a technique that involves the use of contrast fluids through the different absorption of cutaneous melanoma metastases [10]. In the presence of a tumor, the absorption of fluoro-deoxyglucose has a direct correlation with the rate of its proliferation and therefore with the degree of its malignancy. Alternative screening methods are the multifrequency electrical impedance [11] and the Raman spectra [12].

The biological characteristics of a healthy tissue are different from those ones of a diseased tissue. Using the Raman spectra, this different way of diffracting the light coming from a laser beam is exploited, obtaining useful information on the physical structure of the lesion. The spread of wearable devices as well as the enhancement of the smatphones cameras create the conditions for being able to realize applications for auto-screening potentially integrable in platform solutions, through which the patient can interact with specialists aiming at rapid and punctual diagnoses [13]. The increasing costs of health suggest to Governance bodies the adoption of dematerialized health models that can provide decision and responsibility processes, encouraging predictive and preventive analysis [14].

3.2 Image pre-processing

The presence of potential noise linked both to the instrumentation with which the image was acquired, and to the presence of undesired artefacts such as hair and dermatoscopic gel may imply inaccuracies in the diagnosis phase. The objective of the pre-processing phase is to optimize image quality: this phase has a strategic role and is decisive for the entire classification process. The primary task consists in separating the lesion area from the healthy skin. This operation results to be quite difficult as dermoscopic images often present irregular lighting, frames, gelatinous material and intrinsic skin features that make the identification of a lesion, for example trought edges detection methods, a difficult task. Therefore, anything that could affect the image needs to be localized and then removed, masked or replaced. Many techniques have been used in literature in order to perform this task, like image resizing, masking, removal of artefacts, transformation of images into various color spaces. Another important element that commonly penalizes the analysis of dermatoscopic images is the presence of hairs. Recent literature proposes various approaches to optimize images, which involve the use of morphological operators [15], or the detection of curvilinear structures [16] or envisaging a phase of image restructuration through painting [17].

Unfortunately, most of these techniques often leave behind undesirable blurring, invalidating an effective use of the color image. Zhou et al. proposed an algorithm able to carry out an accurate detection of curvilinear artefacts while preserving the lesion structure [18]. The presented approach is very effective for removing the dermatoscopic misuring stick and the hair, but on the other hand it requires high computational requirements. The pre-processing phase of an image can require several operations such as: enhancement, restoration and hair removal (Figure 5).

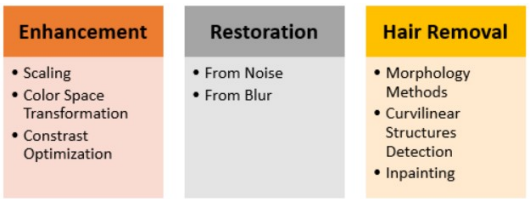


Figure 5: Fundamental steps and techniques for pre-processing

Among the various tools to remove unwanted artefacts, the filters are the most immediate ones. Several Computer Vision Systems dedicated to dermatology adopt filters by applying them directly to grayscale photos or on individual color channels in the case of color images [19, 20, 21].

Generally, for each type of unwanted artefacts, a specific strategy involving the use of dedicated solutions is adopted. Many effective methods use color space transformation to optimize both the quantization of color and of contrast, passing from RGB images to different color spaces for enhancing the variation of the color on the edges and inside the lesion.

The pre-processing phase follows antithetical goals. On the one hand, in order to contain the computational complexity, the objective is to identify a minimal number of colors to be used in image analysis. On the other hand, insufficient contrast would make edge detection very difficult. A solution adopted in many circumstances involves the use of histograms, with which it is possible to perform stretching operations to expand the pixel values in the interval [0, 255], or equalization to obtain a uniform distribution of the pixel values. The important task of appropriately enhancing the contrast between the image area where the lesion is contained and the background consisting of healthy skin can be obtained by compensating for irregular lighting through the use of monomorphic filters and high-pass filters. Adaptive and recursive weighted median filter are used to remove air bubbles and dermoscopic gels [22]. An interesting solution involves local processing carried out on a portion of pixels (generally 3x3), assigning the median to the central pixel. Using these techniques, it is possible to eliminate the noise by containing the alterations that may arise in the original image. The strategies referred to, as well as the other solutions presented in the literature, have as their final objective that of being able to favor the segmentation of the image and the extraction of functional features to the classification process. The strategies to be adopted depend on the nature of the considered images and of the impurities that must be managed [23].

3.3 Image segmentation

The analysis and identification of a lesion can be performed directly by specialists who rely on their experience or by using software solutions for image segmentation, such as IMAGEJ or EDISON (Edge Detection and Image Segmentation system). In fact, in melanoma detection, the main task of the segmentation process is to separate the skin lesion from the background consisting of healthy skin (Figure 6), in order to analyze color, structure, shape, texture and information related to vascularization of the lesion.

Figure 6: (a) dermatoscopic image (b) segmented lesion

The choice of the color space in which it is best to represent images is important, as it determines the rule of the various color channels. Consequently, it will be important to choose the appropriate segmentation algorithm that takes into account the modes of representation chosen for the images. Without the pretension to be exhaustive, some types of segmentation methods in this field are the following:

1. threshold-based, including the Otsu method, those that use the maximum value of the entropy and those that use histograms, such as the independent histogram search algorithm (IHP) [24];

2. region-based, including region growing and watershed segmentation;

3. pixel-based (aimed at labeling the pixel), both supervised methods such as the Support Vector Machine, unsupervised methods such as fuzzy c-means, and those with reinforcement such as artificial neural networks;

4. methods based on customized models, both adaptive and representable through deep multilevel networks.

3.4 Feature extraction

The goal of the segmentation process is to identify the region of the image that contains the lesion by discarding the pixels which are part of the background consisting of healthy skin. In this way the extraction and selection of the most appropriate features to be used for the melanoma detection by artificial vision systems will be favored. In particular, feature extraction implies the enhancement of some specific measures on portions of images: this fundamental step is not often sufficiently discussed and detailed in many researches. On the other hand, a recent study of the relevant features for artificial vision systems in melanoma detection can be found in [6]. As for the extraction of features, many different approaches have been proposed of the statistical and/or filtering type. Among the adopted methods to be mentioned, we cite the principal component analysis, useful for determining the geometric and the shape features, the co-occurrenze matrix [25], useful for the evaluation of texture features, the Fourier spectrum [18], the Gaussian kernels [26], and the transformation of the wavelet package [27], useful for performing image optimization and the extraction of boundaries aimed at reducing data redundancy [28]. In addition, many automated decision support systems implement the feature extraction by exploiting the traditional diagnostic analysis protocols such as the ABCDE system [25], the 7-point checklist [29], the 3-point checklist [30], the model analysis [31] and the Menzies method [32].

4 Numerical experimentations

In this section a preliminary numerical comparison between two different approaches (Support Vector Machine and Multiple Instance Learning) is reported, using color and color/texture features.

This study starts from some considerations related to the works [7, 33, 34]. In [33] the authors analyzed the role played by the color and the texture features, showing empirically that using only the color features outperforms the use of the texture features: very good results were obtained by means of different type of classifiers on a image data set drawn from the Ph^2 database [8], containing 200 melanocytic lesions images (80 common nevi, 80 atypical nevi and 40 melanomas). All the images of the database were selected on the basis of their quality, resolution and dermoscopic features. Good results were obtained also in [7] on the same data set Ph^2,

Table 1: 5-fold cross-validation

	Color			Color and texture		
	SVM	SVM-RBF	MIL-RL	SVM	SVM-RBF	MIL-RL
Correctness (%)	52.50	57.50	71.00	60.50	69.50	74.00
Sensitivity (%)	39.28	60.37	59.21	63.52	71.07	70.59
Specificity (%)	67.29	56.09	83.46	57.81	68.67	77.21
F-score (%)	45.74	58.53	65.99	61.11	69.54	72.10
CPU time (secs)	3.00	0.08	2.52	2.67	0.42	2.98

Table 2: 10-fold cross-validation

	Color			Color and texture		
	SVM	SVM-RBF	MIL-RL	SVM	SVM-RBF	MIL-RL
Correctness (%)	47.50	56.50	71.00	54.50	68.00	72.00
Sensitivity (%)	37.19	61.55	59.63	59.91	70.98	67.30
Specificity (%)	61.75	54.23	83.97	49.62	65.30	78.16
F-score (%)	41.83	57.93	65.14	56.27	68.22	69.54
CPU time (secs)	2.93	0.02	3.26	3.31	0.01	3.82

Table 3: Leave-One-Out validation

	Color			Color and texture		
	SVM	SVM-RBF	MIL-RL	SVM	SVM-RBF	MIL-RL
Correctness (%)	48.50	57.50	71.50	55.00	67.50	69.00
Sensitivity (%)	33.00	64.00	57.00	59.00	69.00	63.00
Specificity (%)	64.00	51.00	86.00	51.00	66.00	75.00
F-score (%)	39.05	60.09	66.67	56.73	67.98	67.02
CPU time (secs)	4.19	0.02	4.44	3.49	0.01	5.50

by applying Multiple Instance Learning techniques and using only color features.

On the other hand, the objective in [34] was to select the most important image features usable in melanoma detection. In this case, differently from [33] where a high quality data set was adopted, the authors performed their experimentations on a data set constituted by plain photographs publicly available from two online databases https://www.dermquest.coml and http://www.dermins.netl, obtaining very good results in terms of sensitivity (96.77%) and, consequently, in terms of F-score (90.91%). The key point of this work consisted in a very accurate pre-processing and segmentation steps.

On the contrary, the objective of our experimentation consists in simply evaluating the classification results, obtainable on 200 images (100 melanomas and 100 common nevi) drawn from the same database used in [34], but without performing any pre-processing step aimed at cleaning up the images from possible noises [23, 35]. This choice is motivated by the necessity to investigate on the possibility to create fast self-diagnosis systems for accessible skin lesions.

In particular, our experiments were performed by using three configurations of the data set, corresponding respectively to the 5-fold cross-validation, the 10-fold cross-validation and the leave-one-out validation.

For each configuration, initially only the color (RGB) features were used, adding, only successively, those ones related to the texture by means of the co-occurence matrix.

The following classification methods, each of them implemented in Matlab (version R2017b), were adopted:

- Support Vector Machine with linear kernel (SVM)[36];
- Support Vector Machine with RBF kernel (SVM-RBF) [36];
- MIL-RL [37].

Some words are to be spent about the MIL-RL method, proposed in [37]. Such a method belongs the class of the Multiple Instance Learning (MIL) algorithms. The MIL paradigm (see for example [38, 39]) is a relatively new approach for classification problems which fits very well on image classification, as shown in the recent works [40, 41] and, in particular, also for medical images and video analysis [42]. It differs from the classical supervised classification approaches, since it is aimed at classifying sets of items: such sets are called *bags* and the items inside the sets are called *instances*. The main characteristic of a MIL algorithm is that, in the learning phase, only the labels of the bags are known, while the labels of the

instances inside the bags are unknown. The MIL paradigm applies very well to image classification since in general an image (bag) is classified on the basis of some portions (instances) of it.

About the MIL-RL code, the same Matlab implementation adopted in [7] were used, while, for the SVM approaches with both linear and RBF kernels, the *fitcsvm* and *predict* subroutines for training the classifier and for computing the testing correctness, respectively, were utilized. Such subroutines are included in the Matlab optimization package.

The results of our experimentations are reported in Tables 1, 2 and 3, in terms of accuracy (testing correctness), sensitivity, specificity, F-score and CPU time. For each of these parameters and for each table, the best result was underlined. It is worth noting that the sensitivity is in general more important than the specificity because it measures the capability to identify positive patients. This is in fact taken into account in the F-score value.

Looking at the tables, it is possible to observe that using both color and texture features improves the results, with respect to using only color features. Moreover the better results are generally obtained in corresponding to the MIL-RL algorithm, whose F-score values are however comparable with those ones obtained by using the SVM with RBF kernel. On the other hand using the SVM-RBF is really faster than using the SVM with linear kernel or the MIL-RL approach.

5 Conclusions

In this work some numerical experimentations were performed to classify a data set constituted by plain photos of different sizes, some of which are low resolution images and others are blurred or with hairs. Although in [34] it has been shown that using some pre-processing methodologies on the images improves the classification performances, in our experiments we preferred do not use such techniques. In this way it was intended to operate in the worst conditions in order to compare the classification performances obtainable by using, on one hand, only color features (as proposed in [40, 41, 43]) and, on the other hand, color and texture features. The fact that the used data set consists of plain photographs, instead of high-quality dermatoscopic images such as those ones constituting the Ph^2 database, justifies the obtained classification performance.

Adopting additional sets of features together with more sophisticated algorithms providing nonlinear separation surfaces, such as in [44], opens up new possible scenarios for the implementation of efficient solutions, useful to provide a further support to specialists and, on the other hand, to have preliminary self diagnoses, since the diffusion of smartphones allows the possibility to take photos of one's own skin injuries. The use of medical solutions [45], as well as the possibility of creating Apps, BOTs and Educational Games [46], designed for different classes of population, are possible scenarios of interest.

References

[1] R. L. Siegel, K. D. Miller, S. A. Fedewa, D. J. Ahnen, R. G. Meester, A. Barzi, and A. Jemal, "Colorectal cancer statistics, 2017," *CA: a cancer journal for clinicians*, vol. 67, no. 3, pp. 177–193, 2017.

[2] R. Pariser and D. Pariser, "Primary care physicians' errors in handling cutaneous disorders: A prospective survey," *Journal of the American Academy of Dermatology*, vol. 17, no. 2, pp. 239–245, 1987.

[3] R. Sanghera and P. S. Grewal, "Dermatological symptom assessment," in *Patient Assessment in Clinical Pharmacy*, pp. 133–154, Springer, 2019.

[4] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, and M. Delfino, "Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: Comparison of the abcd rule of dermatoscopy and a new 7-point checklist based on pattern analysis," *Archives of Dermatology*, vol. 134, no. 12, pp. 1563–1570, 1998.

[5] S. Menzies, C. Ingvar, and W. McCarthy, "A sensitivity and specificity analysis of the surface microscopy features of invasive melanoma," *Melanoma Research*, vol. 6, no. 1, pp. 55–62, 1996.

[6] E. Vocaturo, E. Zumpano, and P. Veltri, "Features for melanoma lesions characterization in computer vision systems," in *2018 9th International Conference on Information, Intelligence, Systems and Applications, IISA 2018*, 2019.

[7] A. Astorino, A. Fuduli, P. Veltri, and E. Vocaturo, "Melanoma detection by means of multiple instance learning," *Interdisciplinary Sciences: Computational Life Sciences, DOI: 10.1007/s12539-019-00341-y*, in press, 2019.

[8] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. S. Marcal, and J. Rozeira, "Ph^2 - A dermoscopic image database for research and benchmarking," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5437–5440, 2013.

[9] G. Anastasi, S. Capitani, M. Carnazza, S. Cinti, O. Cremona, R. De Caro, R. Donato, V. Ferrario, L. Fonzi, A. Franzi, *et al.*, "Trattato di anatomia umana," 2010.

[10] C. Pleiss, J. Risse, H.-J. Biersack, and H. Bender, "Role of fdg-pet in the assessment of survival prognosis in melanoma," *Cancer Biotherapy and Radiopharmaceuticals*, vol. 22, no. 6, pp. 740–747, 2007.

[11] P. ÅAberg, I. Nicander, J. Hansson, P. Geladi, U. Holmgren, and S. Ollmar, "Skin cancer identification using multifrequency electrical impedance - a potential screening tool," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 12, pp. 2097–2102, 2004.

[12] X. Feng, A. Moy, H. Nguyen, Y. Zhang, M. Fox, K. Sebastian, J. Reichenberg, M. Markey, and J. Tunnell, "Raman spectroscopy reveals biophysical markers in skin cancer surgical margins," vol. 10490, 2018.

[13] T. Maier, D. Kulichova, K. Schotten, R. Astrid, T. Ruzicka, C. Berking, and A. Udrea, "Accuracy of a smartphone application using fractal image analysis of pigmented moles compared to clinical diagnosis and histological result," *Journal of the European Academy of Dermatology and Venereology*, vol. 29, no. 4, pp. 663–667, 2015.

[14] E. Vocaturo and P. Veltri, "On the use of networks in biomedicine," *Procedia Computer Science*, vol. 110, pp. 498–503, 2017.

[15] P. Schmid, "Segmentation of digitized dermatoscopic images by two-dimensional color clustering," *IEEE Transactions on Medical Imaging*, vol. 18, no. 2, pp. 164–171, 1999.

[16] M. Fleming, C. Steger, J. Zhang, J. Gao, A. Cognetta, L. Pollak, and C. Dyer, "Techniques for a structural analysis of dermatoscopic imagery," *Computerized Medical Imaging and Graphics*, vol. 22, no. 5, pp. 375–389, 1998.

[17] P. Wighton, T. Lee, and M. Atkins, "Dermascopic hair disocclusion using inpainting," vol. 6914, 2008.

[18] H. Zhou, G. Schaefer, A. Sadka, and M. Celebi, "Anisotropic mean shift based fuzzy c-means segmentation of dermoscopy images," *IEEE Journal on Selected Topics in Signal Processing*, vol. 3, no. 1, pp. 26–34, 2009.

[19] H. W. Lim, H. Honigsmann, and J. L. Hawk, *Photodermatology*. CRC Press, 2007.

[20] I. Maglogiannis, S. Pavlopoulos, and D. Koutsouris, "An integrated computer supported acquisition, handling, and characterization system for pigmented skin lesions in dermatological images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 9, no. 1, pp. 86–98, 2005.

[21] M. Anantha, R. Moss, and W. Stoecker, "Detection of pigment network in dermatoscopy images using texture analysis," *Computerized Medical Imaging and Graphics*, vol. 28, no. 5, pp. 225–234, 2004.

[22] A. Dehghani Tafti and E. Mirsadeghi, "A novel adaptive recursive median filter in image noise reduction based on using the entropy," pp. 520–523, 2013.

[23] E. Vocaturo, E. Zumpano, and P. Veltri, "Image pre-processing in computer vision systems for melanoma detection," in *IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018, Madrid, Spain, December 3-6, 2018*, pp. 2117–2124, 2018.

[24] D. Gómez, C. Butakoff, B. Ersbøll, and W. Stoecker, "Independent histogram pursuit for segmentation of skin lesions," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 1, pp. 157–161, 2008.

[25] N. Abbasi, H. Shaw, D. Rigel, R. Friedman, W. McCarthy, I. Osman, A. Kopf, and D. Polsky, "Early diagnosis of cutaneous melanoma: Revisiting the abcd criteria," *Journal of the American Medical Association*, vol. 292, no. 22, pp. 2771–2776, 2004.

[26] G. Surówka and K. Grzesiak-Kopeć, "Different learning paradigms for the classification of melanoid skin lesions using wavelets," in *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, pp. 3136–3139, 2007.

[27] C. Lee and D. Landgrebe, "Decision boundary feature extraction for neural networks," *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 75–83, 1997.

[28] I. Maglogiannis, E. Zafiropoulos, and C. Kyranoudis, "Intelligent segmentation and classification of pigmented skin lesions in dermatological images," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3955 LNAI, pp. 214–223, 2006.

[29] M. Healsmith, J. Bourke, J. Osborne, and R. Graham-Brown, "An evaluation of the revised seven-point checklist for the early diagnosis of cutaneous malignant melanoma," *British Journal of Dermatology*, vol. 130, no. 1, pp. 48–50, 1994.

[30] H. Soyer, G. Argenziano, I. Zalaudek, R. Corona, F. Sera, R. Talamini, F. Barbato, A. Baroni, L. Cicale, A. Di Stefani, P. Farro, L. Rossiello, E. Ruocco, and S. Chimenti, "Three-point checklist of dermoscopy: A new screening method for early detection of melanoma," *Dermatology*, vol. 208, no. 1, pp. 27–31, 2004.

[31] A. Steiner, H. Pehamberger, and K. Wolff, "In vivo epiluminescence microscopy of pigmented skin lesions. ii. diagnosis of small pigmented skin lesions and early detection of malignant melanoma," *Journal of the American Academy of Dermatology*, vol. 17, no. 4, pp. 584–591, 1987.

[32] S. W. Menzies, K. A. Crotty, C. Ingvar, and W. H. McCarthy, *An atlas of surface microscopy of pigmented skin lesions: dermoscopy*. McGraw-Hill Sydney, Australia, 2003.

[33] C. Barata, M. Ruela, M. Francisco, T. Mendonca, and J. Marques, "Two systems for the detection of melanomas in dermoscopy images using texture and color features," *IEEE Systems Journal*, vol. 8, no. 3, pp. 965–979, 2014.

[34] S. Mustafa and A. Kimura, "A svm-based diagnosis of melanoma using only useful image features," in *2018 International Workshop on Advanced Image Technology, IWAIT 2018*, pp. 1–4, 2018.

[35] E. Vocaturo, E. Zumpano, and P. Veltri, "On the usefulness of pre-processing step in melanoma detection using multiple instance learning," in *13th International Conference on Flexible Query Answering Systems (FQAS) 2019, in press*, 2019.

[36] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013.

[37] A. Astorino, A. Fuduli, and M. Gaudioso, "A Lagrangian relaxation approach for binary multiple instance classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, pp. 2662–2671, 2019.

[38] J. Amores, "Multiple instance classification: review, taxonomy and comparative study," *Artificial Intelligence*, vol. 201, pp. 81–105, 2013.

[39] M.-A. Carbonneau, V. Cheplygina, E. Granger, and G. Gagnon, "Multiple instance learning: A survey of problem characteristics and applications," *Pattern Recognition*, vol. 77, pp. 329–353, 2018.

[40] A. Astorino, A. Fuduli, M. Gaudioso, and E. Vocaturo, "A multiple instance learning algorithm for color images classification," in *Proceedings of the 22Nd International Database Engineering & Applications Symposium*, IDEAS 2018, (New York, NY, USA), pp. 262–266, ACM, 2018.

[41] A. Astorino, A. Fuduli, M. Gaudioso, and E. Vocaturo, "Multiple instance learning algorithm for medical image classification," in *Proceedings of the 27th Italian Symposium on Advanced Database Systems, Castiglione della Pescaia (Grosseto), Italy, June 16-19, 2019*, 2019.

[42] G. Quellec, G. Cazuguel, B. Cochener, and M. Lamard, "Multiple-instance learning for medical image and video analysis," *IEEE Reviews in Biomedical Engineering*, vol. 10, pp. 213–234, 2017.

[43] A. Astorino, A. Fuduli, P. Veltri, and E. Vocaturo, "On a recent algorithm for multiple instance learning. Preliminary applications in image classification," in *Proceedings - 2017 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2017*, vol. 2017-January, pp. 1615–1619, 2017.

[44] M. Gaudioso, G. Giallombardo, G. Miglionico, and E. Vocaturo, "Classification in the multiple instance learning framework via spherical separation," *Soft Computing, DOI: 10.1007/s00500-019-04255-1*, in press, 2019.

[45] E. Zumpano, P. Iaquinta, L. Caroprese, G. Cascini, F. Dattola, P. Franco, M. Iusi, P. Veltri, and E. Vocaturo, "Simpatico 3d: A medical information system for diagnostic procedures," in *IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018, Madrid, Spain, December 3-6, 2018*, pp. 2125–2128.

[46] E. Vocaturo, E. Zumpano, L. Caroprese, S. M. Pagliuso, and D. Lappano, "Educational games for cultural heritage," in *Proceedings of 1st International Workshop on Visual Pattern Extraction and Recognition for Cultural Heritage Understanding co-located with 15th Italian Research Conference on Digital Libraries (IRCDL 2019), CNR Area in Pisa, Italy, January 30, 2019*, pp. 95–106, 2019.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 5, 23-29 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Application of Open-Source Optimization Library "Extremum" to the Synthesis of Feedback Control of a Satellite

Andrei Panteleev*, Valentin Panovskiy

Moscow Aviation Institute (National Research University), department "Mathematics and Cybernetics", Moscow, Russia

ARTICLE INFO

Article history:
Received: 28 May, 2019
Accepted: 20 August, 2019
Online: 06 September, 2019

Keywords:
Open-source
Optimization algorithm
Optimal control

ABSTRACT

Current work demonstrates how open-source optimization library "Extremum" (OSOL Extremum) can be used to build feedback controller of a satellite. Proposed software was developed to as an attempt to eliminate current problems that are present in scientific area: black-box effect (i.e. there is no opportunity to explore source code, modify it, or simply verify), no code reuse (i.e. implemented procedures are accessible only within software that includes it), limitated application of modern optimization algorithms (i.e. number of optimization algorithms increases but most of them were verified only on synthetic tests). All of them lead to so-called reproducibility crisis.

1 Introduction

This paper is an extension of work originally presented in 2018 IV International Conference on Information Technologies in Engineering Education (Inforino) [1].

Optimization is a field of mathematics that is widely applied in engineering, e.g. optimal component design and control theory [2, 3, 4]. Traditionally, optimization algorithms are divided into two groups:

- analytic – these are determined algorithms which efficiency and convergence properties are proved; their main disadvantage is relatively poor performance which reduces the scope of applied problems that can be solved by them [5],
- (meta)heuristic – on the contrary, this group of algorithms has problems with the convergence of results: it is not guaranteed that an appropriate solution will be found, but these algorithms are much more computationally efficient and practically verified [6].

The current work consists of several parts:

- *Prerequisites – explains reasons which led to the OSOL Extremum development;*
- *Description of OSOL Extremum – provides information about software structure and technology stack;*
- *Synthesis of Feedback Control of a Satellite via the OSOL Extremum – describes mathematical task formulation;*
- *Application Results – demonstrates particular results of application of the developed software;*
- *Conclusion – summarizes the results and provides information about further research.*

2 Prerequisites

In order to understand why the development of a new software was performed it is necessary to describe its background. It will be explained in details in the following subsections.

2.1 Black-Boxing

This is the double-edged sword of software development. On the one hand, it allows one to use given tools without necessity to implement desired algorithm, but on the other - it deprives the opportunity to modify code the way you need it for the exact task. In other words, black-boxing makes all your software tools not as flexible as they can be. Also, you must be totally confident in the quality of the software that you use, because you have no opportunity to check confusing moments.

*Corresponding Author: Andrei Panteleev, Moscow Aviation Institute (National Research University), department "Mathematics and Cybernetics", Moscow, Russia, avpanteleev@inbox.ru

2.2 Code Reuse

Generally, academic scientists and software developers work differently. Mainly, scientists are more concerned to solve the given problem without any care about reuse of the scripts and software that they have developed. Programmers, conversely, think a bit differently: they use different programming concepts (e.g. object-oriented programming paradigm) in order to maximize the efficiency of written code in future. That gives an opportunity to create reusable tools speeding up further investigations.

2.3 Reproducibility Crisis

Last but not least, reproducibility crisis is general problem for the science. A lot of researches that are described in scientific articles cannot be verified or at least repeated. That happens because of the restrictions on the article size (i.e. there is no enough space to provide step-by-step description or to explain every moment in detail) and displaced informational focus (i.e. scientists are more concerned to present application results not the specification moments).

2.4 Section Summary

Mentioned problems lead to various limits in overall development speed. This coerces scientist to solve tasks that are out of his professional scope. One possible way to weaken effects of described problems is to develop open-source libraries that offer wide capabilities of API (application programming interface) including unified standard for algorithm creation, general testing environment, and huge spectrum of applied tasks.

Available open-source packages (e.g. python's scikit and scipy packages, R's specialized packages for optimization) partially solve described problems but mostly provide out-of-box tools. Also none of them fully support interval optimization algorithms and they are hardly connected to external programs. That is why it is important to develop library that can be used as set of building blocks for any desired algorithm.

3 Description of OSOL Extremum

This section provides information about OSOL Extremum presenting its main ideas and features. Source code of the library can be found on GitHub site [7].

3.1 Technology Stack

Software part of the developing library consists of modules that are developed using two program languages: Scala (main logic) and Python (for plotting, scripting, and non-math routines).

Scala belongs to multi-paradigm programming languages which support both object-oriented and functional features [8]. Object-oriented paradigm gives opportunity to develop inheritance hierarchy that simplifies code reuse a lot. Functional paradigm is often treated as more "native" to mathematicians because its main statements: treating functions as basic objects, data immutability, pure functions, and etc.). Mainly, these statements aimed to provide code stability decreasing places of potential code mistakes. Also. It should be noted that Scala runs on JVM platform, so it is easily maintained and distributed.

Python is a well-known language which is widely used by scientists of different specialization because of its relative simplicity and generous number of various packages (e.g. NumPy, SciPy, Matplotlib, and etc.) [9].

3.2 Library Structure

OSOL Extremum consists of several parts:

- implementation of basic required mathematical operations and objects, e.g. vector operations, function transformations, random number generators, and statistics calculation,
- general template for optimization algorithms that can be extended by other researches to implement desired method,
- block of implemented optimization algorithms,
- block of different tasks (including basic optimization tasks and complicated optimal control synthesis problems).

3.3 Current State and Future Goals

Current version of OSOL Extremum supports the following features:

- definition of basic optimization algorithm in a form of abstract template that supports wide range of running instructions, e.g. termination by max time or iterations and logging instructions,
- bunch of classical and modern optimization algorithms (including Random Search, Simulated Annealing, Cat Swarm Optimization, and Harmony Search) [5, 6],
- tools for the interaction with Simulink models.

In future it is planned to: implement more optimization algorithms, add support of interval optimization algorithms, add support of other modelling systems (e.g. Wolfram Mathematica), add support of other computational cores.

4 Synthesis of Feedback Control of a Satellite via the OSOL Extremum

This section will provide information how OSOL Extremum can be used to solve applied problem in the field of control theory.

4.1 Stage 1: Open-Loop Control Determination

Problem Definition. During the first stage it is required to solve the problem of optimal open-loop control determination.

Let's consider that the behavior of an object that is described by the following differential equation [10]:

$$\dot{x}(t) = f(t, x(t), u(t)), \tag{1}$$

where $t \in T = [t_0; t_1]$ is continuous time, T - system operational time, $x \in \mathbb{R}^n$ - system state vector, $u \in U \subset \mathbb{R}^q$ - control vector, U - set of admissible control values, $f(t,x,u) = (f_1(t,x,u), ..., f_n(t,x,u))^{\mathrm{T}}$ - continuous vector-function.

Initial state is given:

$$x(t_0) = x_0, \tag{2}$$

terminal state $x(t_1)$ should satisfy the following conditions:

$$\Gamma_i(x(t_1)) = 0, i = 1, ..., l, \tag{3}$$

where $0 \le l \le n$, $\Gamma_i(x), i = 1, ..., l$ are continuously differentiable, system of vectors $\left\{\frac{\partial \Gamma_i(x)}{\partial x_1}, ..., \frac{\partial \Gamma_i(x)}{\partial x_n}\right\}, i = 1, ..., l$ is linearly independent $\forall x \in \mathbb{R}^n$.

The only information which is used to form control is continuous time t. Hence, we consider open-loop control.

Set of admissible processes $D(t_0, x_0)$ is defined as a set of pairs $d = (x(\cdot), u(\cdot))$ which include trajectory $x(\cdot)$ and admissible control $u(\cdot)$ where $\forall t \in T : u(t) \in U$ that satisfy equations (1), (2), and (3).

On the set of admissible processes we consider cost functional

$$I(d) = \int_{t_0}^{t_1} f(t, x(t), u(t))\, dt + F(x(t_1)), \tag{4}$$

where $f^0(t,x,u)$ and $F(x)$ are given continuous functions.

In order to transform the given problem into the problem without terminal constraints we can use penalty function:

$$\bar{I}(d) = I(d) + \sum_{i=1}^{l} E_i(x(t_1), \varepsilon_i, p_i), \tag{5}$$

where

$$E_i(x, \varepsilon_i, p_i) = \begin{cases} (p_i \cdot \Gamma_i(x))^2 \text{ if } |\Gamma_i(x)| > \varepsilon_i \\ |\Gamma_i(x)| \text{ otherwise,} \end{cases} \tag{6}$$

are error functions, p_i and ε_i, $i = 1, ..., l$ are penalty parameters and maximum allowed errors accordingly.

The task is to find such a pair $d^* = (x^*(\cdot), u^*(\cdot))$ that

$$d^* = \mathrm{Arg} \min_{d \in D(t_0, x_0)} I(d). \tag{7}$$

To solve task defined by equation (7) it is suggested to propose parametric form of control and convert initial problem to a problem of parametric optimization. The control, which is obtained by such a procedure, is suboptimal.

Application of Optimization Algorithms. To apply optimization algorithms it is required to parameterize controls in any suitable vector form. Proposed approach was tested using several parameterization schemes:

- polynomial control: $(a_0, a_1, a_2, ..., a_s)^{\mathrm{T}} \Rightarrow u(t) = a_0 + a_1 \cdot t + a_2 \cdot t^2 + ... + a_s \cdot t^s$,
- piecewise linear control: $(a_0, a_1, a_2, ..., a_s)^{\mathrm{T}} \Rightarrow u(t) = \frac{\tau_{i+1} - t}{\tau_{i+1} - \tau_i} \cdot a_i + \frac{t - \tau_i}{\tau_{i+1} - \tau_i} \cdot a_{i+1}$, where $\tau_i = t_0 + i \cdot \frac{t_1 - t_0}{s}$ and $\tau_i \le t < \tau_{i+1}$.

The value of the cost functional is calculated by the numerical methods using Euler, Euler-Cauchy or Runge-Kutta formulas [11, 12].

To obtain numerical results it is proposed to use two techniques:

- simply generate polynomial control (this technique will be called direct),
- start with determination of controls in polynomial form (with few degrees of freedom) and then use obtained solution as a seed for optimization procedure with piecewise linear controls (this technique will be called sequential).

Transformation, which is used during sequential technique, can be easily done by calculation of values of polynomial control in nodes of uniform grid (demonstrated on Figure 1).

Described techniques differ in this way: direct requires less time to obtain control, sequential requires more time (because of two parts) but potentially generates more efficient controls (due to bigger number of degrees of freedom).

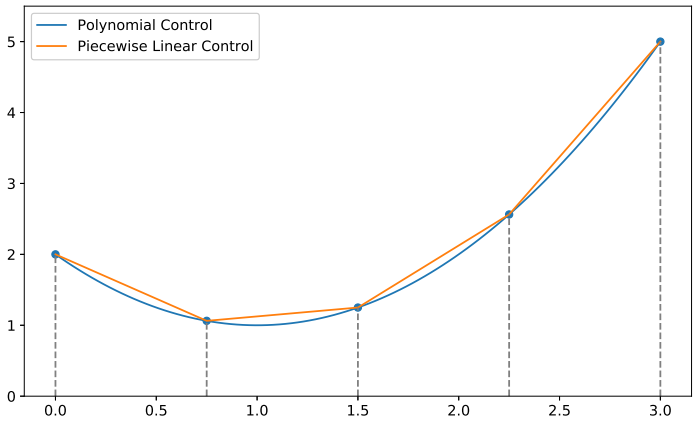


Figure 1: Conversion of polynomial control to piecewise linear one

Thus, initial task is transformed to the determination of optimal vector $(a_0, a_1, a_2, ..., a_s)^{\mathrm{T}}$. As the final result of the first stage set of suboptimal controls $\tilde{u}^{*,k}, k = 1, ..., K$ will be found, i.e. $\tilde{u}^{*,k}$ corresponds to initial state vector $x^k(t_0)$ (K is a number of totally observed initial conditions).

4.2 *Stage 2: Controller Construction*

Final form of the controller will be found during the final stage in a form of regressor. Current block consists of data preparation part and actually regressor construction.

Data Processing. During current phase initial data will be transformed and modified. There are various data processing procedures [13], e.g.

- **Cleaning**. This process ensures that only valid data will be used. In our case it is important to include this phase because on previous stage we use metaheuristic optimization algorithms which are not guaranteed to retrieve the best possible solution. Thus, we will eliminate from further consideration those controls which do not provide sufficient accuracy in the means of the final state error defined by equation (8).
- **Imputation** is widely used in machine learning. Mostly it is present in cases when there is a possibility of data loss. Considered case is determined so there is no need to include it.

- **Transformation**. During this phase data, which was previously cleaned, is transformed from raw format (in our case it is set of suboptimal controls) into format that is suitable for machine learning procedures. We will transform data to a table format.
- **Normalization** is required to standardize obtain data, e.g. nullify its mean values, normalize standard deviation, etc.

In terms of cleaning it is proposed to use only those controls which correspond to trajectories with relatively small terminal constraints errors, e.g.

$$\frac{1}{l} \cdot \sum_{i=1}^{l} |\Gamma_i (x(t_1))| \leq b, \tag{8}$$

where b is a bias value that can be treated as an acceptable level of average cumulative error.

Having set of suboptimal controls $\left\{\tilde{u}^{*,k}\right\}_{k=1}^{K}$ it is possible to find corresponding set of trajectories $\left\{\tilde{x}^{*,k}\right\}_{k=1}^{K}$ from equation (1). This information is sufficient to build initial dataset that will be considered as a base. Specifically, it is proposed to build dataset in a form of Table 1.

Table 1: Initial dataset template

j(index)	t	x_1	$\cdots$	x_n	u_1	$\cdots$	u_q
1	$t\Big\vert_{j=1}$	$x_1\Big\vert_{j=1}$	$\cdots$	$x_n\Big\vert_{j=1}$	$u_1\Big\vert_{j=1}$	$\cdots$	$u_q\Big\vert_{j=1}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$

Such a table can be built for any selected control and then all obtained tables can be concatenated. Thus, we have K different trajectories from each of which we sample in N discrete time moments. Finally, it results into the following bounds for j index: $j = 1, ..., K \cdot N$.

Regressor Construction. Regression is a classical example of so-called supervised learning [14]. In a considered scope of optimal control determination we can treat controls (derived from optimization procedures during the first stage) as an ideal output. Time, and state values can be treated as the input. In other words, desired feedback controller can be treated in this way:

$$\hat{u}_i \left(t\Big\vert_{j=j^*}, x_1\Big\vert_{j=j^*}, ..., x_n\Big\vert_{j=j^*} \right) = u_i\Big\vert_{j=j^*}, \tag{9}$$

where $i = 1, ..., q$ and j^* - possible index.

From equation (9) it is obvious how to generate control for previously observed values but it is not clear what to do with new values that will arise in future. That is why machine learning is proposed to be used here - because of its ability to generalize information basing on finite set of observations [15, 14].

In current work it is proposed to build regressor using well known and widely applied Random Forests [16], which are already implemented in Python open-source library Scikit-Learn [17].

4.3 Section Summary

The core component of the described approach is application of optimization algorithms which can be realized via the OSOL Extremum. Following sections will demonstrate how the proposed approach can be applied to a real control problem.

5 Application Results

5.1 Description of Dynamic Model

To demonstrate the proposed approach the problem of satellite rotational motion dampening is considered [18]. After the conversion to non-dimensional variables system of differential equations, which describes solid body motion relative to centre of mass, will have the following form:

$$\begin{cases} \dot{p} = \frac{u_1}{6}, \\ \dot{q} = u_2 - 0.2 \cdot p \cdot r, \\ \dot{r} = 0.2 \cdot (u_3 + p \cdot q). \end{cases} \tag{10}$$

Control vector constraints are represented via intervals $\forall t \in [0; 1]$: $U = [-500; 500] \times ... \times [-500; 500]$.

Cost functional is given by the formula:

$$I(d) = \int_0^1 (|u_1(t)| + |u_2(t)| + |u_3(t)|)\, dt. \tag{11}$$

Initial values for p_0, q_0, r_0 will be taken from sets $\{-25, -20, ..., -5, 0, 5, ..., 20, 25\}$ (for direct optimization procedure) and $\{21, 22, 23, 24, 25\}$ (for sequential optimization procedure) which results into 11^3 and 5^3 possible combinations accordingly.

Terminal constraints are the same for any starting configuration:

$$p(1) = q(1) = r(1) = 0. \tag{12}$$

Parameters of error functions are also the same for any starting configuration:

$$p_i = 10^3, \varepsilon_i = 0.1, i = 1, 2, 3. \tag{13}$$

Thus, modified functional cost will have the following from:

$$\begin{aligned} \bar{I}(d) = \int_0^1 (|u_1(t)| + |u_2(t)| + |u_3(t)|)dt + \\ + E_1\left(p(t_1), 0.1, 10^3\right) + \\ + E_2\left(q(t_1), 0.1, 10^3\right) + \\ + E_3\left(r(t_1), 0.1, 10^3\right). \end{aligned} \tag{14}$$

5.2 Intermediate Results

For All graphics in this and following section horizontal axis is associated with dimensionless time, vertical axis provides information for target variable (also in dimensionless units).

All results in the current section were provided assuming $s = 2$ for polynomial control and $s = 10$ for piecewise linear.

Figure 2 demonstrates the obtained results during the optimization stage (for sequential technique).

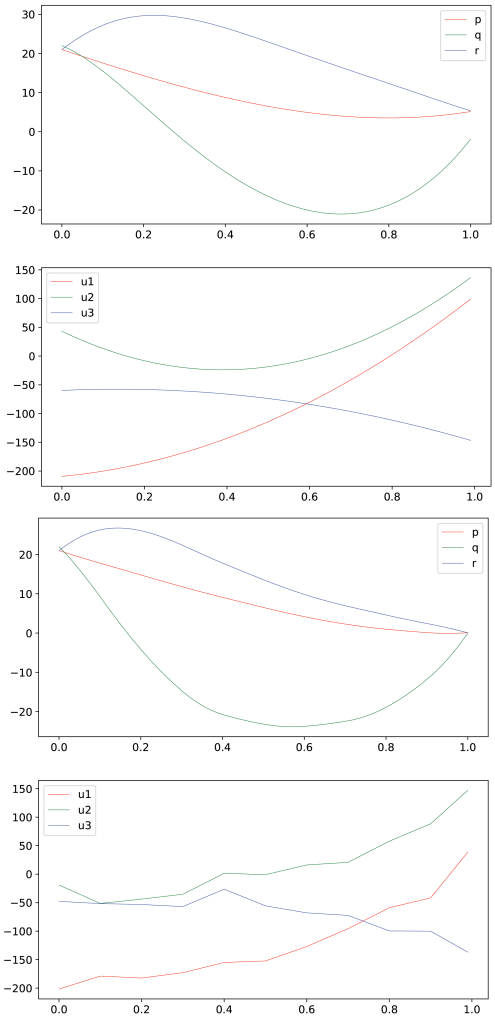


Figure 2: Results of sequential optimization procedure for initial conditions $p(0) = 21, q(0) = 22, r(0) = 21$

From the plot it is clear that during the first stage (generation of polynomial controls) some of them have corresponding trajectories with relatively big errors for terminal state. But it is successfully minimized during the construction of the piecewise polynomial control.

Better visual description of obtained controls can be obtained via bunch graphics demonstrated on Figure 3.

It should be noted that all columns will be scaled (except of the control columns), i.e. transformed to have maximum value equal to 1 and minimum - to 0.

5.3 *Final Results*

The following section will contain information about obtained regressors. Initial values are chosen in order not to duplicate previously used initial states. Each received trajectory was sampled $N = 101$ times in moments $t_i = 0.01 \cdot i$.

Direct technique. Example of application is demonstrated on Figure 4 and Figure 5.

Initial conditions: $p_0 = 24, q_0 = 16, r_0 = 16$.

Achieved terminal state: $p(t_1) = -0.067, q(t_1) = -0.065, r(t_1) = -0.094$.

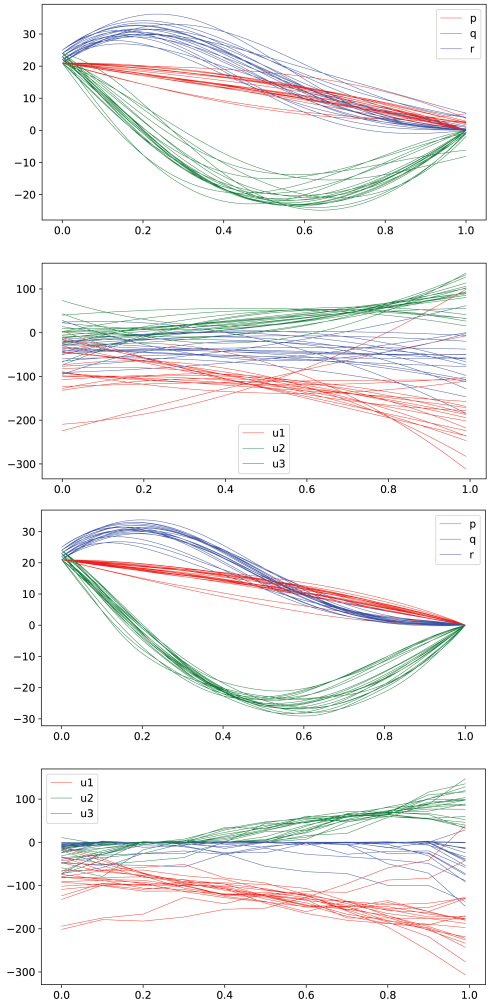


Figure 3: Bunches of trajectories and controls (for sequential optimization procedure)

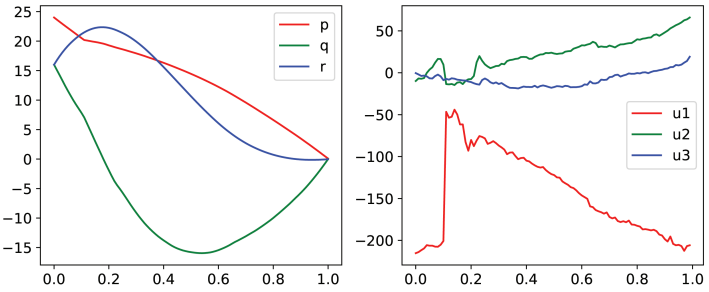


Figure 4: Example 1: trajectories and controls

Initial conditions: $p_0 = -9, q_0 = 13, r_0 = -24$.

Achieved terminal state: $p(t_1) = 0.096, q(t_1) = -0.008, r(t_1) = -0.019$.

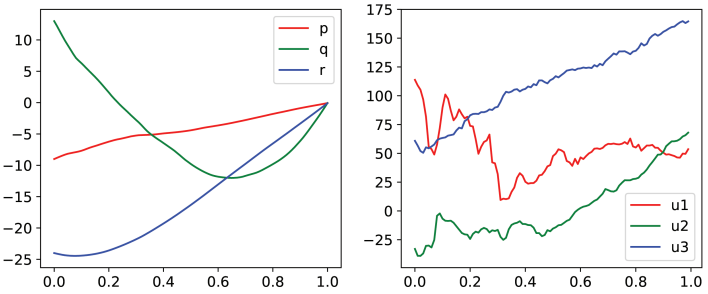


Figure 5: Example 2: trajectories and controls

Sequential technique. Example of application is demonstrated on Figure 6 and Figure 7.

Initial conditions: $p_0 = 24.7, q_0 = 23.3, r_0 = 21.1$.

Achieved terminal state: $p(t_1) = 0.101, q(t_1) = -0.063, r(t_1) = -0.002$.

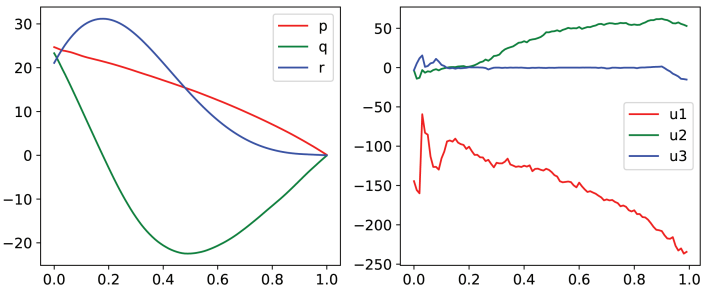


Figure 6: Example 3: trajectories and controls

Initial conditions: $p_0 = 21.1, q_0 = 21.2, r_0 = 21.3$.

Achieved terminal state: $p(t_1) = 0.09, q(t_1) = -0.067, r(t_1) = 0.058$.

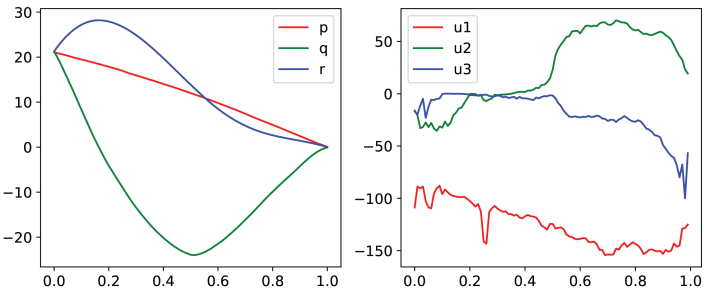


Figure 7: Example 4: trajectories and controls

From observed graphics it is evident that the proposed approach has a great potential to solve the problem of feedback control synthesis. Disadvantage of the proposed approach is that obtained controls are not smooth functions. Such particularity is treated as a disadvantage because it complicates the hardware implementation of the generated control and lowers durability of control mechanism due to the necessity of high-amplitude control adjustments.

6 Conclusion

Main results of the current work are the following:

- efficiency of the developed OSOL Extremum software complex was proved based on obtained solutions for tasks of open-loop control determination,
- proposed approach was tested on an applied task of satellite rotational motion dampening.

Further research is planned to be conducted in several different directions:

- find the way to make controls, that are generated by regressors, have smoother form,
- determine losses in efficiency in the terms of cost functional compared to other computational approaches (including analytic ones),
- find the way how to conduct proper comparative analysis and provide detailed comparison report with the existing software complexes in the terms of accuracy, computational efficiency, and algorithm coverage.

References

[1] Panteleev A.V., Panovskiy V. "2018 IV International Conference on Information Technologies in Engineering Education (Inforino)" in 2018 IV International Conference on Information Technologies in Engineering Education (Inforino), Moscow, Russia, 2018. https://doi.org/10.1109/INFORINO.2018.8581745

[2] Panteleev A.V., Letova T.A., Pomazueva E.A. "Parametric Design of Optimal in Average Fractional-Order PID Controller in Flight Control Problem". Automation and Remote Control, **79**(1): 153 – 166, 2018. https://doi.org/10.1134/S0005117918010137

[3] Kuo Y.-L., Wu T.-L. "Open-loop and closed-loop attitude dynamics and controls of miniature spacecraft using pseudowheels" Computers and Mathematics with Applications, **64**: 1282 – 1290, 2012. https://doi.org/10.1016/j.camwa.2012.03.072

[4] Panteleev A.V., Panovskiy V.N., Korotkova T.I. "Interval explosion search algorithm and its application to hypersonic aircraft modelling and motion optimization problems". Bulletin of the South Ural State University, Series: Mathematical Modelling, Programming and Computer Software, **9**: 55 – 67, 2016. https://doi.org/10.14529/mmp160305

[5] Bazaraa M.S., Sherali H.D., Shetty C.M. Nonlinear programming. Theory and algorithms, New York, John Wiley & Sons, 2006.

[6] Gendreau M., and Potvin J.-S. Handbook of Metaheuristics, New York, Springer, 2010.

[7] GitHub: Open-Source Optimization Library - Extremum. https://github.com/wol4aravio/OSOL.Extremum

[8] Odersky M., Spoon L., Venners B. Programming in Scala, Walnut Creek, Artima Press, 2016.

[9] Lutz M. Learning Python, Sebastopol, O'Reilly Media, 2013.

[10] Panteleev A., Bortakovskiy A. Control Theory in Examples and Problems (Teorija upravlenija v primerah i zadachah), Moscow, Viskaya Shkola, 2003.

[11] Panteleev A., Metlitskaya D. "An application of genetic algorithms with binary and real coding for approximate synthesis of suboptimal control in deterministic systems", Autom. Remote Control, **72**:11, 2328 – 2338, 2011. https://doi.org/10.1134/S0005117911110075

[12] Panteleev A., Metlitskaya D. "Using the method of artificial immune systems to seek the suboptimal program control of deterministic systems", Autom. Remote Control, **75**:11, 1922 – 1935, 2014. https://doi.org/10.1134/S0005117914110034

[13] VanderPlas J. Python Data Science Handbook, Sebastopol, CA: O'Reilly, 2017.

[14] Herbrich R. Graepel T. Machine Learning: An Algorithmic Perspective, Boca Raton, FL, CRC Press, 2015.

[15] Goodfellow I., Bengio Y., Courville A. Deep Learning, Cambridge, MA, MIT Press, 2016.

[16] Breiman L. "Random Forests". Machine Learning, **45**(1): 5 – 32, 2001. https://doi.org/10.1023/A:1010933404324

[17] Hackeling G. Mastering Machine Learning with scikit-learn, Birmingham, Packt Publishing, 2014.

[18] Krylov I. "Numerical solution of the problem of the optimal stabilization of an artificial satellite". USSR Computational Mathematics and Mathematical Physics, **8**(1): 284 – 291, 1968. http://dx.doi.org/10.1016/0041-5553(68)90021-9

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 5, 30-38 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Fully Homomorphic Encryption Scheme Based On Complex Numbers

Khalil Hariss*[1,2], Maroun Chamoun[1], Abed Ellatif Samhat[2]

[1] *Université Saint Joseph, ESIB, CIMTI, Mar Roukoz, Lebanon*

[2] *Lebanese University, Faculty of Engineering, CRSI, Hadath, Lebanon*

ARTICLE INFO

Article history:
Received: 15 June, 2019
Accepted: 20 August, 2019
Online: 06 September, 2019

Keywords:
Complex Numbers
Cloud Systems
Fully Homomorphic
Asymmetric Scheme
Bootstrapping
Approximate GCD

ABSTRACT

In this paper, we present a new Somewhat Homomorphic Encryption (SHE) scheme using computation over complex numbers. We then use Bootstrapping technique to make the scheme Fully Homomorphic (FH) and supports unbounded number of circuit depth. In addition to its homomorphic properties and security level, a main characteristic of the proposed new scheme is its simplicity as it is merely based on addition and multiplication operations over complex numbers. The new scheme is implemented under Python using SAGEMath library and evaluated. Then a crypt-analysis based on Approximate GCD problem is done. A comparison with the BGV, a well known Fully Homomorphic Encryption (FHE) scheme, shows that this new scheme is an efficient homomorphic encryption scheme.

1 Introduction

An encryption scheme is said to be FH, if it allows to perform addition and multiplication operations explicitly over the cipher-texts while performing implicitly the same operations over the plain-texts. This new type of ciphering is very required in cloud systems mainly when users treat the cloud as a third untrusted party. With Homomorphic Encryption (HE) the cloud is able to process over encrypted storage and the privacy of the user is preserved at the cloud side. In Figure. 1, we show the case of a mobile user sending an encrypted query to the cloud using HE, thus the cloud is able to process encrypted query over encrypted data to return an encrypted answer. The user can do the decryption.

The notion of homomorphism was first introduced by Rivest, Adleman and Dertouzo [1] as privacy homomorphism after the invention of RSA crypto-system [2]. The basic RSA crypto-system is a multiplicative homomorphic scheme that allows to perform multiplication operations over the cipher-texts. Give, for example, RSA public key, $p_k = (N, e)$ and cipher $c_i = (m_i)^e mod(N)$ for a plain-text m_i. It is simple to demonstrate that $\prod_i c_i = (\prod_i m_i)^e mod(N)$. Several works followed the RSA scheme. A state of art of HE is given in [3] – [6]. Some of the known homomorphic schemes are: Paillier crypto-system [7, 8], Domingo Ferrer crypto-system [9, 10] which is a HE scheme based on polynomial calculations. The Enhanced MORE [11] is another FHE based on matrix calculation and the NOHE [12] scheme is a lightweight FHE that profits from the simplicity and the homomorphic behavior of logic *NOT*. The most valued work in this domain was given by Craig Gentry in his Ph.D. thesis in 2009 [13, 14]. The work of Gentry was inspired from lattice based cryptography. Gentry first introduced a SHE scheme that supports a bounded number of operations over the cipher-texts. Then he developed a new refresh mechanism called Bootstrapping [15] – [19] to make the scheme FH and supports unbounded operations. An important FHE scheme is the BGV crypto-system developed by Brakerski, Gentry and Vaikuntanathan in [20, 21]. Since BGV scheme is also Somewhat Homomorphic (SH), Modulus Switching (MS) is a technique used with BGV to extend the circuit's depth during the evaluation procedure.

In this paper, our main motivation is to add a value in homomorphic encryption by designing a new simple and robust FHE based on simple complex numbers operations . We first build a new asymmetric SHE complex-based scheme, then we apply Gentry refresh mechanism (Bootstrapping). In comparison with BGV which is lattice based encryption, our scheme is simply based on complex addition and multiplication operations. In addition, Bootstrapping supports unbounded number of circuit depth in our scheme, while MS extends the BGV evaluation to a limited number.

The rest of the paper is organized as follows: in section 2, we present the basic concept of HE with the properties that can lead to a HE scheme. In section 3, we introduce our new FHE scheme built using complex numbers theory. In section 4 we consider our implementations for the new Complex-based scheme and the

*Corresponding Author: Khalil Hariss, Université Saint Joseph, khalil.hariss@net.usj.edu.lb

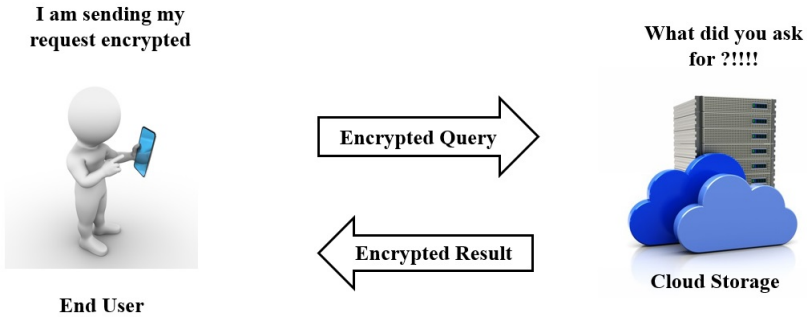


Figure 1: Cloud Scenarion Under Homomorphic Encryption

BGV scheme under Python using SAGEMath Library, followed by a crypt-analysis of our new scheme based on the Approximate GCD problem [22, 23, 24]. And finally the conclusion, comparison between the two schemes and future works are listed in section 5.

2 Homomorphic Encryption

In this work, we focus on asymmetric schemes. Consider an asymmetric scheme α, where p_k is the public key, s_k is the secret key, c_i is the ciphered text of a plain-text m_i. A homomorphic scheme has 3 different basic functions that are: $KeyGen_\alpha$ that generates (p_k, s_k), $Encrypt_\alpha(p_k, m_i)$ that outputs cipher-text c_i, $Decrypt_\alpha(s_k, c_i)$ that outputs the plain-text m_i.

2.1 Evaluation Function

In addition to the three basic functions listed above, a homomorphic scheme has also an evaluation function defined by $Evaluate_\alpha$ that takes as input the public key p_k, a circuit L and a tuple of cipher-texts $C = (c_1, c_2, c_3, ..., c_t)$, cipher of the input plain-texts vector $(m_1, m_2, ..., m_t)$. The evaluation function outputs a cipher-text Ψ given by:

$$\Psi = Evaluate_\alpha(p_k, L, C) \quad (1)$$

The scheme is homomorphic if we have: $\Psi = Encrypt_\alpha(p_k, L(m_1, m_2, ..., m_t))$.

2.2 Homomorphic Properties

When the circuit L performs a certain operation over the encrypted data as written in 1, the cloud is computing a predefined Boolean function f that can be written in a polynomial form. A polynomial form is a set of addition and multiplication gates. Thus to build a HE scheme we should satisfy these two basic properties:

1. Addition

$$Enc_\alpha(p_k, m_1) + Enc_\alpha(p_k, m_2) = Enc_\alpha(p_k, m_1 + m_2) \quad (2)$$

2. Multiplication

$$Enc_\alpha(p_k, m_1) \times Enc_\alpha(p_k, m_2) = Enc_\alpha(p_k, m_1 \times m_2) \quad (3)$$

where m_1 and m_2 are two plain-texts in $\{0, 1\}$ and $Enc_\alpha(p_k, m_i)$ is the encryption function.

3 Homomorphic Complex Scheme

In this section, we build our scheme using complex numbers. We first list the security parameters and then we detail our proposal.

3.1 Parameters

Based on a security parameter λ, different other parameters are generated as listed in [15] to build the new scheme:

1. γ: the bit length of the real and imaginary parts in the public key ($\gamma = \omega(\eta^2 log_2 \lambda)$).
2. ρ: the bit-length of the real and imaginary parts in the noise ($\rho = \omega(log_2(\lambda))$.
3. η: the bit length of the secret key ($\eta \geq \rho.\Theta(\lambda log_2^2(\lambda))$.
4. ρ': extra noise parameter.
5. τ: The number of public keys ($\tau \geq \gamma + \omega(log_2 \lambda)$).

3.2 Somewhat Homomorphic Complex Scheme Construction

The SH complex scheme construction is based on the following steps:

1. Secret key: The secret key s_k is defined by an η bit prime integer p.
2. Public keys: We build a public key bank (PK) formed of τ complex numbers having the following form:
   $PK = \{({p_{k1}}^h + i{p_{k2}}^h) = ((p{q_1}^h + {\epsilon_1}^h) + i(p{q_2}^h + {\epsilon_2}^h), 1 \leq h \leq \tau\}$.
   For each complex public key $({p_{k1}}^h + i{p_{k2}}^h)$, we build different parameters as follow:
   (a) random integer: $q_j^h \longleftarrow Z \cap [0, \frac{2^\eta}{p})$, random noise parameter: ${\epsilon_j}^h \longleftarrow Z \cap (-2^\rho, 2^\rho)$ where $j \in \{1, 2\}$.
   (b) $({p_{k1}}^h, {p_{k2}}^h)$ should have different parities.
   (c) $({\epsilon_1}^h, {\epsilon_2}^h)$ should have the same parity.

3. Encryption Function: The encryption of any bit m is given by the following equation:

$$c = Enc(p_k, m) = ({p_{k1}}^h + i{p_{k2}}^h) \times ({R_1}^h + i{R_2}^h) + {r_1}^h + i{r_2}^h + m \quad (4)$$

where $({p_{k1}}^h + i{p_{k2}}^h)$ is chosen randomly from the public key bank (PK), ${R_1}^h + i{R_2}^h$ is a random complex chosen such that $|{R_1}^h| <<< \frac{p}{6|{\epsilon_1}^h - {\epsilon_2}^h|}$, $|{R_2}^h| <<< \frac{p}{6|{\epsilon_1}^h + {\epsilon_2}^h|}$, ${r_1}^h + i{r_2}^h$ is a random complex chosen between $(-2^{\rho'}, 2^{\rho'})$ such that ${r_1}^h$ and ${r_2}^h$ have the same parity and $|{r_1}^h - {r_2}^h| \ll \frac{p}{6}$(Choosing the intervals of R_1^h, R_2^h, r_1^h, r_2^h is explained in the upcoming decryption section).

4. Decryption Function:
Following 4, we can demonstrate that $c = Enc(p_k, m) = ({p_{k1}}^h + i{p_{k2}}^h) \times ({R_1}^h + i{R_2}^h) + {r_1}^h + i{r_2}^h + m = (pq_1{}^h R_1{}^h - pq_2{}^h R_2{}^h + R_1{}^h \epsilon_1{}^h - R_2{}^h \epsilon_2{}^h + r_1{}^h + m) + i(pq_1{}^h R_2{}^h + pq_2{}^h R_1{}^h + R_2{}^h \epsilon_1{}^h + R_1{}^h \epsilon_2{}^h + r_2{}^h)$.
The decryption of the cipher c is done as follows:

$$\begin{aligned} Intermediate = real(c) - imag(c) = \\ p(q_1{}^h R_1{}^h - q_2{}^h R_2{}^h - q_1{}^h R_2{}^h - q_2{}^h R_1{}^h) + \\ R_1{}^h(\epsilon_1{}^h - \epsilon_2{}^h) - R_2{}^h(\epsilon_1{}^h + \epsilon_2{}^h) + (r_1{}^h - r_2{}^h) + m. \\ m = mod(modNear(Intermediate, p), 2). \end{aligned} \quad (5)$$

Where $modNear(x, p) = y$, such that $y \in (-\frac{p}{2}, +\frac{p}{2})$.
Decryption works as long as:
$\frac{-p}{2} \leq {R_1}^h({\epsilon_1}^h - {\epsilon_2}^h) - {R_2}^h({\epsilon_1}^h + {\epsilon_2}^h) + ({r_1}^h - {r_2}^h) + m \leq \frac{p}{2}$.
In addition $({\epsilon_1}^h - {\epsilon_2}^h)$, $({\epsilon_1}^h + {\epsilon_2}^h)$ and $({r_1}^h - {r_2}^h)$ are even integers.
The decryption condition is satisfied as long as: $|{R_1}^h| <<< \frac{p}{6|{\epsilon_1}^h - {\epsilon_2}^h|}$, $|{R_2}^h| <<< \frac{p}{6|{\epsilon_1}^h + {\epsilon_2}^h|}$, $|{r_1}^h - {r_2}^h| << \frac{p}{6}$ and each pair of $({\epsilon_1}^h, {\epsilon_2}^h)$, $({r_1}^h, {r_2}^h)$ is formed of two integers having the same parity.

3.3 Complex Homomorphic Properties

We show in this section that the proposed scheme satisfies the homomorphic properties. Given two cipher-texts c_1, c_2 respectively for 2 different plain-texts m_1, m_2 encrypted using the complex scheme listed in section 3.2.
$c_1 = ({p_{k1}}^h + i{p_{k2}}^h)({R_1}^h + i{R_2}^h) + {r_1}^h + i{r_2}^h + m_1 = (pq_1{}^h R_1{}^h - pq_2{}^h R_2{}^h + R_1{}^h \epsilon_1{}^h - R_2{}^h \epsilon_2{}^h + r_1{}^h + m_1) + i(pq_1{}^h R_2{}^h + pq_2{}^h R_1{}^h + R_2{}^h \epsilon_1{}^h + R_1{}^h \epsilon_2{}^h + r_2{}^h)$. $c_2 = ({p_{k1}}^t + i{p_{k2}}^t)({R_1}^t + i{R_2}^t) + {r_1}^t + i{r_2}^t + m_2 = (pq_1{}^t R_1{}^t - pq_2{}^t R_2{}^t + R_1{}^t \epsilon_1{}^t - R_2{}^t \epsilon_2{}^t + r_1{}^t + m_2) + i(pq_1{}^t R_2{}^t + pq_2{}^t R_1{}^t + R_2{}^t \epsilon_1{}^t + R_1{}^t \epsilon_2{}^t + r_2{}^t)$.

1. Addition

$$\begin{aligned} c_1 + c_2 = \\ (pq_1{}^h R_1{}^h - pq_2{}^h R_2{}^h + pq_1{}^t R_1{}^t - pq_2{}^t R_2{}^t + R_1{}^h \epsilon_1{}^h \\ - R_2{}^h \epsilon_2{}^h + R_1{}^t \epsilon_1{}^t - R_2{}^t \epsilon_2{}^t + r_1{}^h \\ + r_1{}^t + m_1 + m_2) + i(pq_1{}^h R_2{}^h + pq_2{}^h R_1{}^h + \\ pq_1{}^t R_2{}^t + pq_2{}^t R_1{}^t + R_2{}^h \epsilon_1{}^h + R_1{}^h \epsilon_2{}^h + R_2{}^t \epsilon_1{}^t \\ + R_1{}^t \epsilon_2{}^t + r_2{}^h + r_2{}^t). \end{aligned} \quad (6)$$

Decryption of $c_1 + c_2$ is obtained by calculating $mod(modNear(real(c_1 + c_2) - imag(c_1 + c_2), p), 2)$ having the following form:
$pQ_{add} + R_{add} + m_1 + m_2$ such that $R_{add} = {R_1}^h({\epsilon_1}^h - {\epsilon_2}^h) - {R_2}^h({\epsilon_1}^h + {\epsilon_2}^h) + {R_1}^t({\epsilon_1}^t - {\epsilon_2}^t) - {R_2}^t({\epsilon_1}^t + {\epsilon_2}^t) + ({r_1}^h - {r_2}^h) + ({r_1}^t - {r_2}^t)$.
Decryption works as long as R_{add} is even and $\frac{-p}{2} \leq R_{add} \leq \frac{p}{2}$.

2. Multiplication
Similar to addition, decryption of $c_1 \times c_2$ is obtained by calculating $mod(modNear(real(c_1 \times c_2) - imag(c_1 \times c_2), p), 2)$.
We can demonstrate that $real(c_1 \times c_2) - imag(c_1 \times c_2)$ have the following form:
$pQ_{mult} + R_{mult} + m_1 m_2$ such that $R_{mult} = ({r_1}^t {R_1}^h + m_2 {R_1}^h - {r_2}^t {R_2}^h)({\epsilon_1}^h - {\epsilon_2}^h) - ({r_1}^t {R_2}^h + {r_2}^t {R_1}^h + m_2 {R_2}^h)({\epsilon_1}^h + {\epsilon_2}^h) + ({r_1}^h {R_1}^t + m_1 {R_1}^t - {r_2}^h {R_2}^t)({\epsilon_1}^t - {\epsilon_2}^t) - ({R_1}^t {r_2}^h + {r_1}^h {R_2}^t + m_1 {R_2}^t)({\epsilon_1}^t + {\epsilon_2}^t) + ({R_1}^h {R_1}^t {\epsilon_1}^h - {R_2}^h {R_2}^t {\epsilon_1}^h)({\epsilon_1}^t - {\epsilon_2}^t) + ({R_2}^h {R_2}^t {\epsilon_2}^h - {R_1}^h {R_1}^t {\epsilon_2}^h)({\epsilon_1}^t + {\epsilon_2}^t) - ({R_1}^h {R_2}^t {\epsilon_1}^h + {R_2}^h {R_1}^t {\epsilon_1}^h)({\epsilon_1}^t + {\epsilon_2}^t) - ({R_2}^h {R_1}^t {\epsilon_2}^h + {R_1}^h {R_2}^t {\epsilon_2}^h)({\epsilon_1}^t - {\epsilon_2}^t) + {r_1}^h({r_1}^t - {r_2}^t) - {r_2}^h({r_1}^t + {r_2}^t) + m_2({r_1}^h - {r_2}^h) + m_1({r_1}^t - {r_2}^t)$.
Decryption also works as long as R_{mult} is even and $\frac{-p}{2} \leq R_{mult} \leq \frac{p}{2}$. We can simply demonstrate that R_{add} and R_{mult} are even since each couple of $({r_1}^h, {r_2}^h)$, $({r_1}^t, {r_2}^t)$, $({\epsilon_1}^h, {\epsilon_2}^h)$ and $({\epsilon_1}^t, {\epsilon_2}^t)$ is formed of two integers having the same parity, but having R_{add} and R_{mult} between $\frac{-p}{2}$ and $\frac{p}{2}$ is not always satisfied.
As a conclusion our new complex scheme is SH that supports a bounded number of addition and multiplication operations over the cipher-texts.

3.4 Making the Complex Scheme Fully Homomorphic

To make our complex scheme FH, we apply Bootstrapping [15, 17] in order to reduce the noise level after each operation. Starting from a complex cipher c, the plain-text m can be calculated following this equation $m \leftarrow [c^* - \lfloor \frac{c^*}{p} \rceil]_2$ where $c^* = real(c) - imag(c)$ since this decryption equation is much simpler than Equation. 5. First of all we can use Gentry's transformation to squash the decryption circuit. In this transformation we add to the public key some extra information about the secret key, and use this extra information to

post process the cipher-text c^*. The post processed cipher-text c^* can be decrypted more efficiently than the original cipher-text.

3.4.1 Bootstrapping:

Consider the evaluation procedure given in 1, we use instead of the arithmetic circuit L, the decryption circuit D_α of the SH scheme α. The main concept of Bootstrapping is that we have a cipher-text ψ_1 that encrypts m under p_{k1} that we want to refresh. s_{k1} is the secret key related to the public key p_{k1}. s_{k1} is encrypted under another public key p_{k2}. Let $\overline{s_{k1j}}$ be the encrypted secret bits j of s_{k1} under p_{k2}. Bootstrapping is given by this algorithm:

$$\begin{aligned} & Recrypt(p_{k2}, D_\alpha, < \overline{s_{k1j}} >, \psi_{1j}), \\ & Set \quad \overline{\psi_{1j}} \leftarrow Encrypt(p_{k2}, \psi_{1j}) \\ Output \quad & \psi_2 \leftarrow Evaluate_\alpha(p_{k2}, D_\alpha, \langle \overline{\langle s_{k1} \rangle}, \langle \overline{\psi_{1j}} \rangle \rangle) \end{aligned} \tag{7}$$

In 7, the function $Evaluate_\alpha$ takes as input the bits of s_{k1} and ψ_1, each encrypted under p_{k2} to evaluate homomorphically the decryption circuit D_α. The output ψ_2 is an encryption under p_{k2} of $Decrypt_\epsilon(s_{k1}, \psi_1) = m$. By applying 7, we are removing the error vector associated to the first cipher and adding another error vector. The progress is made as long as the second error vector in ψ_2 is shorter than the primitive in ψ_1.
In Figure. 2, we present the evaluation procedure of a circuit with high depth giving birth to a cipher c with a high noise level.
As given in Figure. 3, for a cipher c consider $D_c(sk) = Decrypt_{sk}(c)$, this operation is called squashing the decryption circuit (will be explained in the upcoming section) and $D_c(.)$ is a low depth polynomial in sk. Bootstrapping consists of evaluating $D_c(.)$ using the encryption of the secret key sk.

3.4.2 Squashing the Decryption Circuit:

Bootstrapping is possible as long as the decryption circuit is an arithmetical circuit of low depth. Squashing is the required transformation in making it possible as listed in[15, 16, 17, 18, 19]. Following Gentry's Squashing technique, we add three different extra parameters κ, θ, Θ that are function of the security parameter λ. $\kappa = \frac{\gamma\eta}{\rho'}$, $\theta = \lambda$, $\Theta = \omega(\kappa.log(\lambda))$. For a secret key $s_k^* = p$ and a public key p_k^* from the original complex homomorphic scheme α, we add to the public key a set $\overrightarrow{Y} = \{y_1, y_2, y_3, ..., y_\Theta\}$ of rational numbers in $[0, 2)$ with κ bits of precision, such that there is a sparse subset $S \subset \{1, 2, 3, ..., \Theta\}$ of size θ with $\sum_{i \in S} y_i \approx \frac{1}{p} (mod 2)$.
The secret key becomes the indicator of the subset S. The scheme of section 3.2 becomes FH complex by first applying the following modifications:

1. KeyGen:
   Generate $s_k^* = p$ and p_k^* as before. set $x_p \leftarrow \lfloor \frac{2^\kappa}{p} \rceil$. Choose randomly a Θ bit vector $\overrightarrow{s}$ with hamming weight θ, $\overrightarrow{s} = \langle s_1, s_2, ..., s_\Theta \rangle$ and let $S = \{i; s_i = 1 \text{ in vector } S\}$. Choose Θ random integers $u_i \in Z \cap [0, 2^{\kappa+1})$, $i = 1, 2, 3, ..., \Theta$ subject to the condition that $\sum_{i \in S} u_i = x_p mod(2^{\kappa+1})$. Set $y_i = \frac{u_i}{2^\kappa}$ and $\overrightarrow{Y} = \{y_1, y_2, ..., y_\Theta\}$. Hence each y_i is a positive number smaller than 2 with κ bits of precision after the binary point such that $[\sum_{i \in S} y_i]_2 = (\frac{1}{p}) - \Delta_p$ for some $\Delta_p < 2^{-\kappa}$.
2. Encrypt and Evaluate Algorithm:
   Generate the cipher-text $c^* = real(c) - imag(c)$ from a complex cipher c. Then for each $i \in \{1, 2, 3, ..., \Theta\}$, set $z_i \leftarrow [c^*.y_i]_2$ keeping only $n = \lceil log\theta \rceil + 3$ bits of precision after the binary point for each z_i. Output both c^* and $\overrightarrow{z} = \langle z_1, z_2, ..., z_\Theta \rangle$
3. Decrypt and Output Algorithm:

$$m \leftarrow [c^* - \lfloor \sum_{i \in S} s_i z_i \rceil]_2 \tag{8}$$

Lemma 1. *For every cipher-text $(c^*, \overrightarrow{z})$ that is generated by evaluating a circuit C, it holds that $\sum_i s_i z_i$ is within $\frac{1}{4}$ of an integer.*

Proof. Fix an arithmetic circuit L, public keys and secret keys generated with respect to the security parameter λ, with $\{y_i\}_{i=1}^\Theta$ the rational numbers in the public key and $\{s_i\}_{i=1}^\Theta$ the secret key bits. If we recall that y_i's were chosen that: $[\sum_i s_i y_i]_2 = (\frac{1}{p}) - \Delta_p$. We fix a circuit L, and t cipher-texts $\{c_i\}_{i=1}^t$ that encrypts the input to L and denote: $c = Evaluate_\epsilon(pk, L, c_1, c_2, .., ..., c_t)$, we need to establish this equation:

$$\lfloor \frac{c^*}{p} \rceil = \lfloor \sum_i s_i z_i \rceil (mod 2) \tag{9}$$

where $c^* = real(c) - imag(c)$ and the z_i's are computed as $[c^*.y_i]_2$ with only $\lceil log\theta + 3 \rceil$ bits of precision after the binary point, hence $[c^*.y_i]_2 = z_i - \Delta_i$ with $|\Delta_i| \leq \frac{1}{16\theta}$.
$[(\frac{c^*}{p}) - \sum s_i z_i]_2 = [(\frac{c^*}{p}) - \sum_i s_i [c^* y_i]_2 + \sum s_i \Delta_i]_2 = [(\frac{c^*}{p}) - c^* [\sum s_i y_i]_2 + \sum s_i \Delta_i]_2 = [(\frac{c^*}{p}) - c^*.(\frac{1}{p} - \Delta_p) + \sum s_i \Delta_i]_2$
$= [c^*.\Delta_p + \sum s_i \Delta_i]_2$. □

We claim that the final quantity inside the brackets has magnitude at most $\frac{1}{8}$. By definition, since c^* is a valid cipher-text output, the value $\frac{c^*}{p}$ is within $\frac{1}{8}$ of an integer. Together all these facts imply the lemma. To establish the claim, we first observe that $|\sum s_i \Delta_i| \leq \theta \times \frac{1}{16\theta}$. Regarding $c^* \Delta_p$, recall that the output c^* is obtained by evaluating the circuit L on the cipher-text c_i as listed in [15], for any polynomial P that implements the circuit L, for any $\alpha \geq 1$, if P's input has magnitude at most $2^{\alpha(\rho'+2)}$, its output magnitude at most $2^{\alpha(\eta-4)}$. In particular, when P's are fresh cipher-texts, which have magnitude at most 2^γ, P's output cipher-text c^* has magnitude at most $2^{\gamma(\eta-4)(\rho'+2)} < 2^{\kappa-4}$. Thus $|c^* \Delta_p| < \frac{1}{16}$.

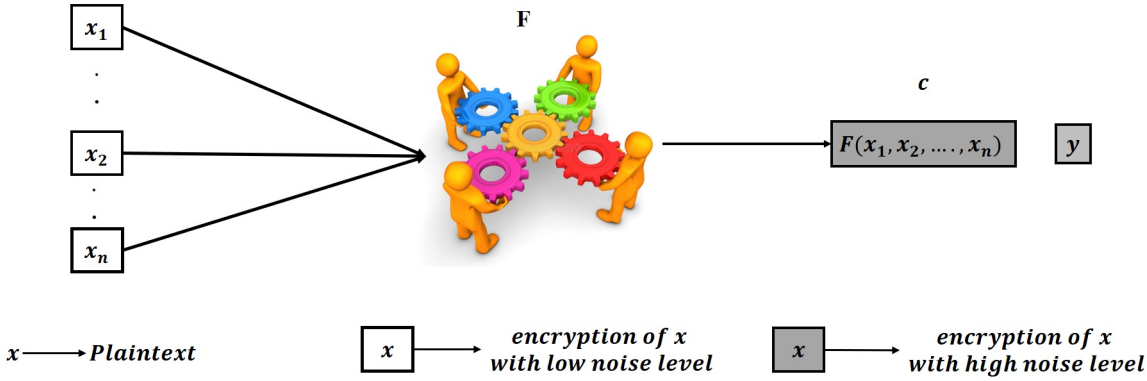


Figure 2: Circuit Evaluation with High Noise

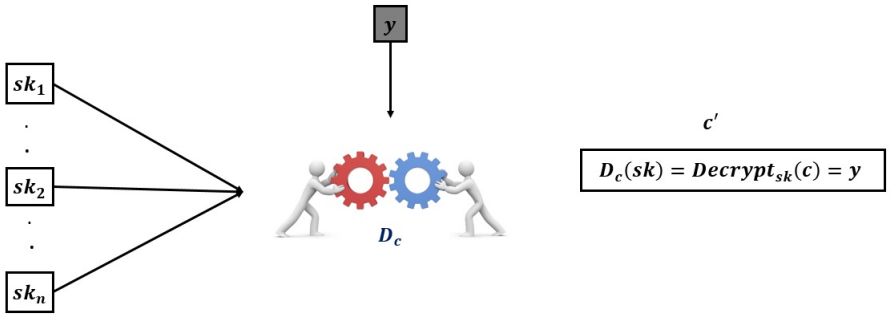


Figure 3: Fresh Cipher Generation

3.4.3 *Practical Implementation of Squashing and Bootstrapping:*

For the squashed decryption circuit given in 8, the implementation of $\sum_{i\in S} s_i z_i$ (such that $z_i = (z_{i1}, z_{i2}, ..., z_{in})$ is a vector of $n = \lceil log\theta \rceil + 3$ bits, where $1 \leq i \leq \Theta$), can be done with lower computational complexity as implemented in [16, 19]. This new implementation is based on dividing the new secret key $\overrightarrow{s}$ in θ boxes of $B = \dfrac{\Theta}{\theta}$ bits each, where each box has a single bit having the value 1. This will lead us to obtain a grade school addition algorithm that requires $O(\theta^2)$ multiplications instead of $O(\Theta.\theta)$. The secret key $\overrightarrow{s}$ is divided into $s_{k,i}$, the i^{th} secret key bit in box k, where $1 \leq k \leq \theta$ and $1 \leq i \leq B$. The resultant equation is:

$$m \leftarrow (c^* - \lfloor \sum_{k=1}^{\theta} (\sum_{i=1}^{B} s_{k,i} z_{k,i}) \rceil) mod(2) \tag{10}$$

We denote that the sum $q_k = \sum_{i=1}^{B} s_{k,i} z_{k,i}$ is obtained by adding B numbers, only one being nonzero. The decryption equation is now:

$$m \leftarrow (c^* - \lfloor \sum_{i=1}^{\theta} (q_k) \rceil) mod(2) \tag{11}$$

Where the q_k's are rational in $[0, 2)$ with n bits of precision after the binary point. Another form of the decryption equation is given by:

$$m = [c^*]_2 \oplus [\lfloor \sum_{i=1}^{\Theta} s_i z_i \rceil]_2 \tag{12}$$

Based on 12, we can deduce that the parity of the plain-text m is related to the parity of the primitive cipher-text c^* and the parity of $[\lfloor \sum_{i=1}^{\Theta} s_i z_i \rceil]_2$.

In order to make this concept much clearer, a simple implementation of bootstrapping, using low values, is explained in the following example. Giving randomly $\Theta = 4$, $\theta = 2$, $B = \dfrac{\Theta}{\theta} = 2$, and a precision bit $n = 1$. The secret key $\overrightarrow{s} = [b_1, b_2, b_3, b_4]$ such that ($b_1 = 1$ and $b_2 = 0$) or ($b_1 = 0$ and $b_2 = 1$), ($b_3 = 1$ and $b_4 = 0$) or ($b_3 = 0$ and $b_4 = 1$) based on the implementation of 10. Initial values for evaluating the decryption circuit using the secret key $\overrightarrow{s}$ without encryption are $s_1 = [s_{11} \quad s_{12}] = [b_1 \quad b_2]$, $s_2 = [s_{21} \quad s_{22}] = [b_3 \quad b_4]$, the z values are taken just as an example $z_1 = [z_{11} \quad z_{12}] = [[10] \quad [01]]$, $z_2 = [z_{21} \quad z_{22}] = [[11] \quad [01]]$. We start by applying the evaluation procedure over the plain-texts by calculating $Sum = \sum_{k=1}^{\theta} \sum_{i=1}^{B} s_{k,i} z_{k,i} = \sum_{k=1}^{2} \sum_{i=1}^{2} s_{k,i} z_{k,i} = (b_1 z_{11} + b_2 z_{12}) + (b_3 z_{21} + b_4 z_{22}) = [b_1 \quad b_2] + [b_3 \quad b_3 + b_4] =$

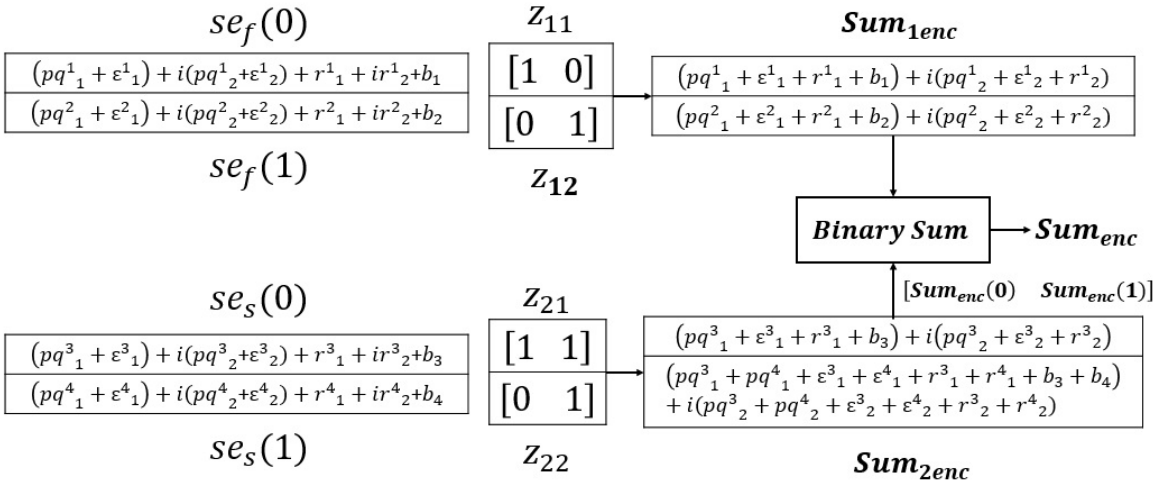


Figure 4: Circuit Evaluation over Cipher-texts.

$(Sum_1) + (Sum_2)$, then we apply the binary summation to calculate $Sum = Sum_1 + Sum_2 = [b_1 + b_3 \quad b_2 + b_3 + b_4 + b_1b_3] = [Sum(0) \quad Sum(1)]$ ($Sum(0) = Sum_1(0) + Sum_2(0)$ and $Sum(1) = Sum_1(1) + Sum_2(1) + Sum_1(0)Sum_2(0)$). As a result, the parity of $\lfloor Sum \rceil$ is given by $b_1 + b_3 + b_2 + b_3 + b_4 + b_1b_3$. Next step is to do the circuit evaluation over the complex cipher-texts. Let $se_f = [se_1 \quad se_2]$ be the cipher of the secret key $s_1 = [b_1 \quad b_2]$ and $se_s = [se_3 \quad se_4]$ be the cipher of the secret key $s_2 = [b_3 \quad b_4]$ using the encryption procedure listed in 4 with random values $(R_1^h + iR_2^h) = 1$.

$se_f = [se_f(0) \quad se_f(1)] = Enc(s_1) = [(pq_1^1 + \epsilon_1^1) + i(pq_2^1 + \epsilon_2^1) + r_1^1 + ir_2^1 + b_1 \quad (pq_1^2 + \epsilon_1^2) + i(pq_2^2 + \epsilon_2^2) + r_1^2 + ir_2^2 + b_2]$.

$$Sum_{enc} = \sum_{k=1}^{\theta}\sum_{i=1}^{B} se_{k,i}z_{k,i} = \sum_{k=1}^{2}\sum_{i=1}^{2} se_{k,i}z_{k,i} = Sum_{1enc} + Sum_{2enc},$$

where $Sum_{1enc} = se_f(0)z_{11} + se_f(1)z_{12}$ and $Sum_{2enc} = se_s(0)z_{21} + se_s(1)z_{22}$. An implementation of circuit evaluation over the cipher-texts is given in Figure.4.

As shown in Figure.4, after applying the binary summation over Sum_{1enc} and Sum_{2enc} we obtain $Sum_{enc} = [Sum_{enc}(0) \quad Sum_{enc}(1)]$ such that $Sum_{enc}(0) = p(Q_{0R} + iQ_{0I}) + (\epsilon_1^1 + \epsilon_1^3) + i(\epsilon_2^1 + \epsilon_2^3) + (r_1^1 + r_1^3) + i(r_2^1 + r_2^3) + \mathbf{b_1} + \mathbf{b_3}$.

$Sum_{enc}(1) = p(Q_{1R}+iQ_{1I})+(\epsilon_1^1\epsilon_1^3-\epsilon_2^1\epsilon_2^3)+(r_1^3\epsilon_1^1+r_1^1\epsilon_1^3-r_2^3\epsilon_2^1-r_2^1\epsilon_2^3)+(r_1^1r_1^3-r_2^1r_2^3)+b_1\epsilon_1^3+b_3\epsilon_1^1+b_1r_1^3+b_3r_1^1+(\epsilon_1^2+\epsilon_1^3+\epsilon_1^4)+(r_1^2+r_1^3+r_1^4)+\mathbf{b_1b_3}+\mathbf{b_2}+\mathbf{b_3}+\mathbf{b_4}+i((\epsilon_1^1\epsilon_2^3+\epsilon_2^1\epsilon_1^3)+(r_2^3\epsilon_1^1+r_1^1\epsilon_2^3+r_2^1\epsilon_1^3+r_1^3\epsilon_2^1)+(r_2^1r_1^3+r_1^1r_2^3)+b_1\epsilon_2^3+b_3\epsilon_2^1+b_1r_2^3+b_3r_2^1+(\epsilon_2^2+\epsilon_2^3+\epsilon_2^4)+(r_2^2+r_2^3+r_2^4))$.

Since $mod(modNear(real(Sum_{enc}(0)) - imag(Sum_{enc}(0)), p), 2) = b_1+b_3$ and $mod(modNear(real(Sum_{enc}(1))-imag(Sum_{enc}(1)), p), 2) = b_1b_3 + b_2 + b_3 + b_4$, as long as the encryption parameters are chosen such that the scheme is SH (i.e Sum_{1enc} and Sum_{2enc} are formed of complex ciphers that can be added and multiplied homomorphically as long as the scheme is SH). Based on all the calculations listed above, a fresh cipher-text of c^* can be written as:

$$fresh_{cipher} = \lfloor Sum_{enc}(0) + Sum_{enc}(1) \rceil + [c^*]_2 + \sum_j (p_{k_1}{}^j + ip_{k_2}{}^j)(R_1^j + iR_2^j) \quad (13)$$

4 Implementation and Security Analysis

In order to validate our work, first we did an implementation of our new Complex-based scheme, then we compared the evaluation of the logic circuit $C = (b_0 \oplus b_1) \bullet (b_2 \oplus b_3)$ using the two schemes: Complex and BGV.

As a brief overview of the BGV [20, 21], it is a FHE scheme that works over bit level and built using Lattice based cryptography. The security of this scheme depends on the hardness of Learning With Error (LWE) introduced by Oded Regev in [25], that relies on the complexity of solving a noisy linear system. Starting from a security parameter λ, we have a secret key $s \in Z_p^{[1,n]}$ and a cipher-text $c \in Z_p^{[n,1]}$. The hardness of LWE resides in taking the lattice dimension: $n \approx poly(\lambda)$ and the ring dimension: $p \approx poly(n)$ following [21, 25]. Homomorphic addition is achieved by simply adding the two cipher-texts, while homomorphic multiplication is done by calculating the Tensor product of the two different cipher-texts which increases the dimension exponentially. Key Switching (KS) is a new technique introduced to reduce the cipher dimension after each homomorphic multiplication. The basic BGV scheme is SH since the noise level will increase with circuit depth, therefore MS is another technique introduced to reduce the noise level after each arithmetic operation and extend to a higher circuit depth.

Finally a crypt-analysis of the new scheme is validated with the GACD attack. All simulations are done under Python with SAGE-Math Library using a machine having the following specifications (CPU: Intel Xeon, E5 – 2630, 2.40 GHZ, 8 CORES, 128 GB RAM).

4.1 Implementations and Results

1. **First Implementation** : First implementation with the new Complex-based scheme is done with three different security layers (**small**: $\lambda = 42$, $\rho = 26$, $\eta = 988$, $\gamma = 147456$, $\Theta = 150$, $\tau = 158$), (**medium**: $\lambda = 52$, $\rho = 41$, $\eta = 1558$, $\gamma = 843033$, $\Theta = 555$, $\tau = 572$), (**large**: $\lambda = 62$, $\rho = 56$, $\eta = 2128$, $\gamma = 4251886$, $\Theta = 2070$, $\tau = 2110$) with extra noise parameter $\rho' = \eta - \rho$, θ the hamming weight of vector $\overrightarrow{s}$ is equal to 15, n the precision after the binary point of each

Table 1: First Implementation Results

Parameters	*KeyBankGen*	*Encrypt*	*Decrypt*
small	0.136439000002 *s*	0.00102999999945 *s*	0.000683999998728 *s*
medium	2.732191 *s*	0.00130100000024 *s*	0.00339099999837 *s*
large	48.049091 *s*	0.00596199999927 *s*	0.0216390000023 *s*

Table 2: Second Implementation Results

Parameters	Complex mean execution time	BGV mean execution time
small	0.4102108 *s*	0.35041905 *s*
medium	2.8407822 *s*	2.42358273 *s*
large	27.40187732 *s*	38.52468052 *s*

z_i is taken as 4. Table. 1 shows in terms of execution time the generation of public key bank of τ complex public keys, 1 bit encryption and decryption.

2. **Second Implementation** : In the second implementation, we did the evaluation procedure of the logic circuit $C = Plain_{output} = (b_0 \oplus b_1) \bullet (b_2 \oplus b_3)$, (i.e. $Cipher_{Output} = ((c_0 + c_1) \times (c_2 + c_3)$, where $c_i = Enc(b_i)$, $for\ 0 \leq i \leq 3$) using the two schemes (Complex and BGV) with the same level of security λ. For BGV the three layers of implementation are (**small**: $\lambda = 42$, $n = 42$, $p \approx O(n^{20})$, $q \approx O(\frac{n^{20}}{2})$), (**medium**: $\lambda = 52$, $n = 149$, $p \approx O(n^6)$, $q \approx O(\frac{n^6}{2})$), (**large**: $\lambda = 62$, $n = 370$, $p \approx O(n^6)$, $q \approx O(\frac{n^6}{2})$). Table.2 shows a comparison in terms of mean execution time between the two implementations for 100 iterations (Evaluation procedure for the Compplex-based scheme is done with Bootstrapping mechanism while with BGV is achieved with KS and MS). In addition, results have shown that noise is efficiently reduced for the two schemes. One can see that for large case, the proposed Complex based-scheme performs better that BGV.

4.2 Security Analysis:

The crypt-analysis of the homomorphic encryption schemes will consider the techniques used to build such schemes. The security of the BGV-lattice based scheme relies on the hardness of solving LWE problems [25]. As for our new complex scheme, it uses mathematical operations and the typical attack in this case is General Approximate Common Divisor (GACD). Given a public key bank $PK = \{x_1, x_2, ..., x_\tau\}$ where $x_j = pq_j + r_j$ for $1 \leq j \leq \tau$, the basic idea of this attack is to reveal the value of the secret key p starting from PK. The secret key p can be revealed using different types of algorithms like the approximate GCD of two numbers discussed in [22], the approximate GCD of many numbers using the SDA algorithm [23] by applying a lattice based attack with *LLL* algorithm. Recently a new improved attack was introduced and implemented in [24], [19]. In our crypt-analysis, we tried to apply this new Approximate GCD attack. Given the public key bank $PK = \{x_j, 1 \leq j \leq \tau\}$, where $q_j \in [0, \frac{2^\gamma}{p})$ and $r_j \in [0, 2^\rho)$ are chosen uniformly and independently at random. The algorithm is as follows: For $j = 1, 2, ..., \tau$, let:

$$y_j = \prod_{i=0}^{2^\rho - 1} (x_j - i) \tag{14}$$

Equation 14 shows clearly that p divides the GCD $g = gcd(y_1, y_2, ..., y_\tau)$. To build this attack and depending on the choice of the (q_j, r_j), we will try to find a certain bound B not much larger than 2^ρ that with a high probability, all the prime factors of g except p are smaller than this bound B. The probability that all the prime factors of g except p are smaller than B is done based on [19]: "For every prime $p \geq B$ other than p, not all the x_j's are congruent to one of $(0, 1, ..., 2^\rho - 1) \quad mod p$". This happens with probability very close to $1 - (\frac{2^\rho}{p})^s$. Hence, the probability that all the prime factors of g except p are smaller than B is essentially given by the following Euler product:

$$P_{s,\rho}(B) = \prod_{p \geq B,\ p \neq p} (1 - \frac{2^{s\rho}}{p^s}) \tag{15}$$

Based on [19] and using the usual prime counting function $\pi(x)$ explained in [26], we can demonstrate that 15 converges to some positive number smaller than 1 and satisfies the following lemma:

Lemma 2. *For any $B > 2^{\rho + \frac{1}{s}}$, we have:*
$1 - P_{s,\rho}(B) < \frac{2s}{s-1} \times \frac{2^{s\rho}}{B^{s-1} log B}$

In our simulation, we picked $B = 2^{\frac{s+1}{s-1}\rho}$ and we got a success probability $P_{s,\rho}(B) > 1 - 2^{-\rho}$. The GACD attack is given by the pseudo code of Algorithm 1.

Algorithm 1 *GACD Attack*

procedure p=$GACD_(\eta,\gamma,\rho)$
  $p \leftarrow random_prime(0, 2^{\eta})$
  $s \leftarrow \rho \frac{s+1}{s-1}$
  $B \leftarrow \lfloor 2^{\rho} \rfloor$
  $Fa \leftarrow Factorial(B)$
  $x \leftarrow p \times random_integer(0, 2^{\gamma-\eta}) + random_integer(0, 2^{\rho})$
  $i \leftarrow 0$
  $g \leftarrow 1$
  while $i < 2^{\rho}$ **do**
    $g \leftarrow g \times (x - i)$
    $i \leftarrow i + 1$
  end while
  $j \leftarrow 1$
  while $j \leq s$ **do**
    $x \leftarrow p \times random_integer(0, 2^{\gamma-\eta}) + random_integer(0, 2^{\rho})$
    $i \leftarrow 0$
    $z \leftarrow 1$
    while $i < 2^{\rho}$ **do**
      $z \leftarrow z \times (x - i)$
      $i \leftarrow i + 1$
    end while
    $g \leftarrow Greatest_divisor_of_gcd(g, z)_prime_to_Fa$
    if $\lfloor log_2(g) \rfloor \leq \eta$ **then**
      Break
    end if
    $j \leftarrow j + 1$
  end while
return g
end procedure

Due to the limited resources of our machine (CPU: Intel Xeon, E5-2630, 2.40 GHZ, 8 CORES, 128 GB RAM), the proposed GACD attack with the security levels related to λ is not feasible since the polynomial y_j given in 14 is of degree 2^{ρ} with coefficients of size γ and requires a memory of size $2^{\rho}\gamma$ bits. The required size of each level is: **small** : 1.125 Terra Byte, **medium** : 210759 Terra Byte, **large** : 34831286272 Terra Byte, while our machine is only 128 GB RAM.

5 Conclusion

In this paper, we profited from the simplicity of complex numbers properties by proposing a new SWE scheme based on complex numbers. We applied Gentry refresh mechanism to make our scheme FH. We then implemented our new scheme with the BGV using SAGEMath library. As a comparison with BGV, a well known FHE scheme, our new scheme is an efficient homomorphic scheme and performs better than BGV in terms of execution for large implementation. In addition, our scheme is simply based on simple complex operations rather than lattice based cryptography (homomorphic complex multiplication is done without dimension expansion rather than Tensor product) and Bootstrapping can support unbounded circuit depth, while MS used with BGV is limited to some circuit depth. Finally a crypt-analysis based on GACD attack is presented. Future work will consider the implementation of the GACD Attack given in section 4.2 with a more powerful machine in order to evaluate the approximate attack time.

Acknowledgment This work has been partially funded with support from the Lebanese University.

References

[1] R. Rivest, L. Adleman, and M. Dertouzos, "On data banks and privacy homomorphisms," in *Foundations on Secure Computation, Academia Press*, 1978, pp. 169–179. [Online]. Available: http://www.oalib.com/references/14708317

[2] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Commun. ACM*, vol. 26, no. 1, pp. 96–99, Jan. 1983. [Online]. Available: http://doi.acm.org/10.1145/357980.358017

[3] P. Martins, L. Sousa, and A. Mariano, "A survey on fully homomorphic encryption: An engineering perspective," *ACM Comput. Surv.*, vol. 50, no. 6, pp. 83:1–83:33, Dec. 2017. [Online]. Available: http://doi.acm.org/10.1145/3124441

[4] L. Zhang, Y. Zheng, and R. Kantoa, "A review of homomorphic encryption and its applications," in *Proceedings of the 9th EAI International Conference on Mobile Multimedia Communications*, ser. MobiMedia '16. ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2016, pp. 97–106. [Online]. Available: http://dl.acm.org/citation.cfm?id=3021385.3021405

[5] C. Aguilar-Melchor, S. Fau, C. Fontaine, G. Gogniat, and R. Sirdey, "Recent advances in homomorphic encryption: A possible future for signal processing in the encrypted domain," *IEEE Signal Processing Magazine*, vol. 30, no. 2, pp. 108–117, March 2013. [Online]. Available: http://doi.acm.org/10.1109/MSP.2012.2230219

[6] S. Fau, R. Sirdey, C. Fontaine, C. Aguilar Melchor, and G. Gogniat, "Towards practical program execution over fully homomorphic encryption schemes," in *2013 Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC-2013)*, Compiègne, France, Oct. 2013, pp. –. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00917061

[7] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Advances in Cryptology — EUROCRYPT '99*, J. Stern, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 223–238. [Online]. Available: https://link.springer.com/chapter/10.1007/3-540-48910-X_16

[8] M. Nassar, A. Erradi, and Q. M. Malluhi, "Paillier's encryption: Implementation and cloud applications," in *2015 International Conference on Applied Research in Computer Science and Engineering (ICAR)*, Oct 2015, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/7338149

[9] J. D. i Ferrer, "A new privacy homomorphism and applications," *Information Processing Letters*, vol. 60, no. 5, pp. 277 – 282, 1996. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020019096001706

[10] J. Domingo-Ferrer, "A provably secure additive and multiplicative privacy homomorphism*," in *Information Security*, A. H. Chan and V. Gligor, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 471–483. [Online]. Available: https://dl.acm.org/citation.cfm?id=744660

[11] K. Hariss, H. Noura, and A. E. Samhat, "Fully enhanced homomorphic encryption algorithm of more approach for real world applications," *Journal of Information Security and Applications*, vol. 34, no. Part 2, pp. 233 – 242, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2214212616303052

[12] K. Hariss, H. Noura, A. E. Samhat, and M. Chamoun, "Design and realization of a fully homomorphic encryption algorithm for cloud applications," in *Risks and Security of Internet and Systems*, N. Cuppens, F. Cuppens, J.-L. Lanet, A. Legay, and J. Garcia-Alfaro, Eds. Cham: Springer International Publishing, 2018, pp. 127–139. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-76687-4_9

[13] C. Gentry, "A fully homomorphic encryption scheme," Ph.D. dissertation, Stanford, CA, USA, 2009, aAI3382729. [Online]. Available: https://dl.acm.org/citation.cfm?id=1834954

[14] G. Craig, "Fully homomorphic encryption using ideal lattices," in *Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing*, ser. STOC '09. New York, NY, USA: ACM, 2009, pp. 169–178. [Online]. Available: http://doi.acm.org/10.1145/1536414.1536440

[15] M. van Dijk, C. Gentry, S. Halevi, and V. Vaikuntanathan, "Fully homomorphic encryption over the integers," in *Advances in Cryptology – EUROCRYPT 2010*, H. Gilbert, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 24–43. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-13190-5_2

[16] C. Gentry and S. Halevi, "Implementing gentry's fully-homomorphic encryption scheme," in *Advances in Cryptology – EUROCRYPT 2011*, K. G. Paterson, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 129–148. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-20465-4_9

[17] C. Gentry, S. Halevi, and N. P. Smart, "Better bootstrapping in fully homomorphic encryption," in *Public Key Cryptography – PKC 2012*, M. Fischlin, J. Buchmann, and M. Manulis, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 1–16. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-30057-8_1

[18] J.-S. Coron, A. Mandal, D. Naccache, and M. Tibouchi, "Fully homomorphic encryption over the integers with shorter public keys," in *Advances in Cryptology – CRYPTO 2011*, P. Rogaway, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 487–504. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-22792-9_28

[19] J.-S. Coron, D. Naccache, and M. Tibouchi, "Public key compression and modulus switching for fully homomorphic encryption over the integers," in *Advances in Cryptology – EUROCRYPT 2012*, D. Pointcheval and T. Johansson, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 446–464. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-29011-4_27

[20] Z. Brakerski and V. Vaikuntanathan, "Fully homomorphic encryption from ring-lwe and security for key dependent messages," in *Proceedings of the 31st Annual Conference on Advances in Cryptology*, ser. CRYPTO'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 505–524. [Online]. Available: http://dl.acm.org/citation.cfm?id=2033036.2033075

[21] Z. Brakerski, C. Gentry, and V. Vaikuntanathan, "(leveled) fully homomorphic encryption without bootstrapping," in *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, ser. ITCS '12. New York, NY, USA: ACM, 2012, pp. 309–325. [Online]. Available: http://doi.acm.org/10.1145/2090236.2090262

[22] N. Howgrave-Graham, "Approximate integer common divisors," in *Cryptography and Lattices*, J. H. Silverman, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 51–66. [Online]. Available: https://link.springer.com/chapter/10.1007/3-540-44670-2_6

[23] J. C. Lagarias, "The computational complexity of simultaneous diophantine approximation problems," *SIAM J. Comput.*, vol. 14, no. 1, pp. 196–209, Feb. 1985. [Online]. Available: http://dx.doi.org/10.1137/0214016

[24] Y. Chen and P. Q. Nguyen, "Faster Algorithms for Approximate Common Divisors: Breaking Fully-Homomorphic-Encryption Challenges over the Integers," in *EUROCRYPT 2012*, ser. Lecture Notes in Computer Science, D. Pointcheval and T. Johansson, Eds., vol. 7237, IACR. Cambridge, United Kingdom: Springer, Apr. 2012, pp. 502–519. [Online]. Available: https://hal.inria.fr/hal-00864374

[25] O. Regev, "On lattices, learning with errors, random linear codes, and cryptography," in *Proceedings of the Thirty-seventh Annual ACM Symposium on Theory of Computing*, ser. STOC '05. New York, NY, USA: ACM, 2005, pp. 84–93. [Online]. Available: http://doi.acm.org/10.1145/1060590.1060603

[26] E. Bach and J. Shallit, *Algorithmic Number Theory*. Cambridge, MA, USA: MIT Press, 1996. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-662-04053-9_2

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 5, 39-45 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Low-Cost and Accurate Computational System for Efficiency Measures over Photovoltaic Arrays

Héctor L. Núñez-Ramírez[*,1], Gloria G. Carvalho-Kassar1[1], Freddy C. Brito-Maestre[2], Yaremi I. Gamboa-Maldonado[3], Luis A. Santos-Avendaño[1], Carlelines Gavidia-Toro[4], Orlando Villarroel-Ramos[3], Dino Di Rosa-Ulloa[4]

[1]*Fundación Centro Nacional de Desarrollo e Investigación en Telecomunicaciones, Dirección de Fotónica, 01060A, Caracas, Venezuela.*

[2]*Universidad Central de Venezuela, Departamento de Comunicaciones, 01050, Caracas, Venezuela.*

[3]*Fundación Centro Nacional de Desarrollo e Investigación en Telecomunicaciones, Dirección de Propagación y Antenas, 01060A, Caracas, Venezuela.*

[4]*Fundación Centro Nacional de Desarrollo e Investigación en Telecomunicaciones, Dirección de Electrónica de Comunicaciones, 01060A, Caracas, Venezuela.*

ARTICLE INFO

Article history:
Received: 30 May, 2019
Accepted: 14 July, 2019
Online: 06 September, 2019

Keywords:
Efficiency Measures
Photovoltaic Arrays
Software
Two-diode model

ABSTRACT

Great efforts have been shown in the literature to improve the efficiency of photovoltaic (PV) systems. In this workpaper, authors present the development of a computational system based on free software that allows to estimate with an accuracy superior to 99.9 percent the performance of the PV array connected to a current inverter block, from the physical models behind. For this purpose, a solar radiation measures was made in the laboratory, and also voltage, current and temperature tests over the whole system, adding a storage interface and serial data transmission port to drive and process the data from a computer. Once the system was tested and calibrated in laboratory, the I-V experimental curves of the PV system were obtained in real time, in order to define the most important parameters: short-circuit current, open-circuit voltage, maximum current and voltage, in a natural environment. Those parameters were the input data to the computational tool based on Scilab, which allow estimate the performance of PV system described, solving the non-linear equations of the physical models used. The performance results were compared with a similar computational tool developed in Matlab, as well as with a commercial solar simulator used by the manufacturer of PV module. A highly accurate computational system based in Scilab was obtained allows to evaluate the efficiency and the filling factor of an array for different temperature and solar irradiance levels in real time.

1 Introduction

Great efforts have been made to improve the efficiency in electric power generation using photovoltaic (PV) systems, since the advantages in comparison with others alternative energy sources, being generally fixed systems, which do not require inputs or fuels for operate, just sunlight, which geographically is available in most cases.

This workpaper is an extensión of work originally presented in IEEE 2018 Open Innovations Conference (OI) [1], as a contribution on developing software to determine the behavior of photovoltaic arrays, e.g. [2, 3], in order to be able to size correctly a deployment in any cases: residential, commercial, industrial and even farming environments. Most of these software require license payment or the source code can not be modified. Even, the most recent and interesting works based in two-diode model simulation [4] – [9] use Matlab to evaluate the performing of photovoltaic arrays. However, this work shows the results of an alternative computational system, based on two-diode model used for polycrystalline silicon PV cells [10]– [13], that determines the performance of a cell under sunlight irradiance variations and hence the relevance of knowing the efficiency of the device under study. To achieve this goal, the behavior of a semiconductor in sunlight conditions have been considered as well as temperature changes, to determine the current density between the terminals of the cell, generally consisting of a n-type and p-type semiconductor materials, which are brought

[*]Héctor L. Núñez-Ramírez, Fundación Cendit - Caracas 01060A, +582128088275, hectornunez2002@gmail.com

togheter in the well-known n-p junction. The current of carriers under the photovoltaic phenomenon are deduced by means of the equations that describe the physical model that lead us to a non-linear equations system, which will be solved by software using different Scilab functions. The coefficients of equations have been determined through the most important parameters of any photovoltaic array: short-circuit current (I_{sc}), open-circuit voltage (V_{oc}), maximum current (I_{max}) and voltage (V_{max}), all of them determined experimentally in Photonic Laboratory of Cendit. Once the equations are solved, the I-V and P-V curves of the array under test are obtained, and they are compared with the results obtained using Matlab and the solar simulator facility from one manufacturer of the module under study. However, the results obtained include the performance of the DC-AC current inverter, which is part of the system, using a complete bridge configuration, PWM controller and LC filtering at the output. The computer system developed shows not only the I-V and P-V curves of the device under test but also shows the efficiency and the filling factor of whole set.

This workpaper was developed in five sections: Physical and Mathematical Modeling to focuses phenomena and equations according with the photovoltaic arrays used; Experimental Setup to describe the procedures performed in the laboratory to obtain the data required; Software Development with the scope of describe the tools used and the programming developed, as well as the use of the obtained experimental data; Results and Conclusions to analyze, compare, discuss and verify the operation of the computational tool presented.

2 Physical and Mathematical Modelling

The physical properties of a commercial photovoltaic array made of polycrystalline silicon was considered in this case, constituted by a n-p junction of semiconductor materials under the physical fundamentals of cells of this type described in [14, 15]. Isolated, each semiconductor material retains its neutral charge. The holes are the majority carriers in the p-type material and their positive charge is compensated by negatively-charged ionized acceptor atoms, while the electrons are the majority carriers in the n-type material where there is a large concentration of free electrons compensated by positively-charged ionized donor atoms. If both semiconductor materials are placed together, a very large difference in electron concentration occurs that causes an hole diffusion current from p-type material towards the n-type material as well as an electron diffusion current from the n-type material through the metallurgical junction to the p-type material, but the region close to the junction become free of mobile charge carriers, reason why this area is denoted as a depletion or space-charge region. The diffusion generates minority carriers in each region. The gradual depletion of the charge carriers generates a charge zone made up of acceptors and donor atoms, which is no longer compensated by the free mobile carriers. As quasi-neutral regions are denoted the regions around of the depletion zone, where the neutral charge is maintained (see figures 1 and 2).

In the depletion region, under equilibrium conditions and darkness, an internal electric field is formed which forces the charge carriers to move in the opposite direction to the concentration gradient, therefore, the diffusion current continues its flow, while the forces act on the charge carriers, compensate each other, then the diffusion and drift (due to the internal electric field created around the junction) currents that originate from the carriers are compensated, reason why the net current density is zero ($I_{drift} = I_{diffusion}$). When an increase in temperature occurs, electron-hole pairs are generated, provided that thermal energy can undo the bonds, increasing the minority carriers within the material, which causes a current density due to thermal-generation.

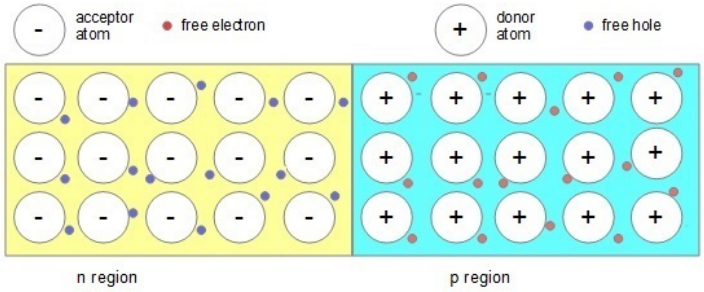


Figure 1: n-p junction (top: before junction).

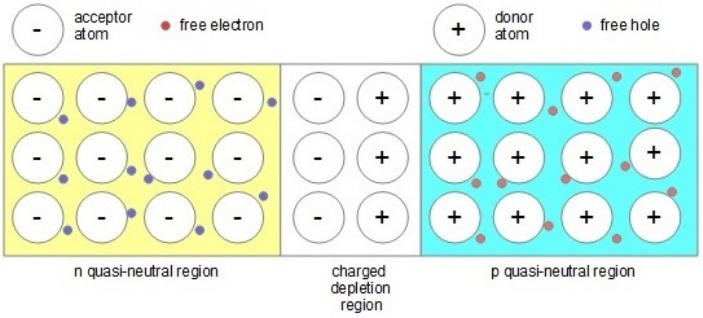


Figure 2: n-p junction (after junction).

The difussion currents (due to minority carriers) into n-type and p-type materials causes recombination currents since the minority carriers recombine with the majority carriers (the electron-hole pairs are undone). The recombination current is compensated by the thermal-generation current ($I_{rec} = I_{gen}$). The recombination strongly determines the performance of the photovoltaic cell, since the current that can be collected by the minority carriers within the semiconductor is reduced.

When the thermal equilibrium is broken, under illumination conditions, due to photons with enough energy (and frequency) to create an electron-hole pair, between the valence and conduction bands of the metallurgical junction, which increases the concentration of the minority carriers, is enabled the flow of holes from the n-type region to the p-type region, and vice versa in the case of the electrons. The flow of the photo-generated carriers determines a photo-generation current density (of intensity I_{ph}), which is added to the current density by thermal generation when the temperature increases within the material.

Based on the above, the behavior of the photovoltaic cell in darkness can be described as a currents balance due to recombinations and thermal generation phenomena, by 1:

$$I_{dark} = I_{01}(e^{qV/kT} - 1) + I_{02}(e^{qV/AkT} - 1) \quad (1)$$

where I_{01} and I_{02} are the saturation current in the quasi-neutral regions and the saturation current in the depletion region, respectively, q is the carrier's charge, k is the Boltzman's constant, T is the temperature, V is the built potential due internal electric field in the junction, A is the ideality factor between both regions. In this case, note that the exponential function is like the Boltzman Approximation, which has been demonstrated in the literature.

On the other hand, under illumination condition, the photogeneration current I_{ph} appears as shown in expression 2:

$$I_{total} = I_{ph} - I_{01}(e^{qV/kT} - 1) - I_{02}(e^{qV/kT} - 1) \tag{2}$$

with:

$$I_{ph} = [I_{sc} - K_i(T - Tr)](\frac{E}{1000\ W/m^2}) \tag{3}$$

where I_{sc} represents the short-circuit current, I_{total} corresponds to carriers net flow between the conducting terminals from the photovoltaic array, K_i is the short-circuit current coefficient, T_r represents the reference temperature and E is the irradiance.

The I-V and P-V curves, current-voltage and power-voltage respectively, can be obtained from equation (2), whose performance is shown in figure 3, where the I_{sc} is pointed at the left extreme when $V = 0$ and V_{oc} in the right extreme at $I_{total} = 0$.

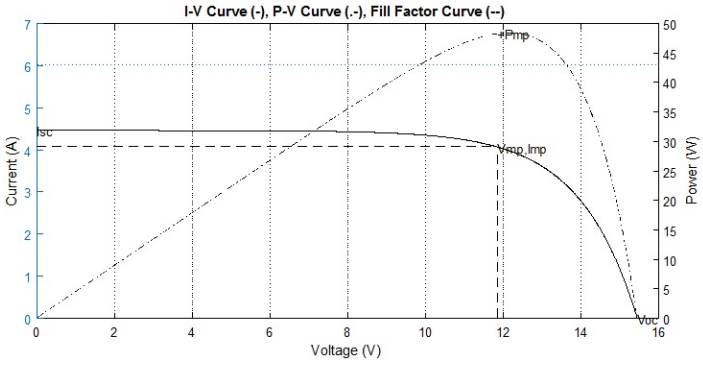


Figure 3: I-V Curve (blue), P-V Curve (red) and Fill Factor (yellow).

When the short-circuit condition is given, I_{max} not could be reached completly, in fact is a little lower due to the high resistivity presents between the ends of the junction, equivalent to a resistance in parallel to the current source and the diodes, denoted as shunt resistance (R_{sh}). In the other side, when the open-circuit condition is given the V_{max} not could be reached due to the ohmic losses in the union of the contacts with the ends of the n-p junction, denoted as resistance in series (R_s) to the current source and diodes.

$$I_{total} = I_{ph} - I_{01}(e^{\frac{q(V-I_{total}Rs)}{kT}} - 1) - I_{02}(e^{\frac{q(V-I_{total}Rs)}{AkT}} - 1) - (V + I_{total}R_s)/R_{sh} \tag{4}$$

The behavior of a single photovoltaic cell is given by the combination in parallel or series of a single photovoltaic cell, modifying expression (4) by the following:

$$I_{total} = N_pI_{ph} - N_pI_{01}(e^{q(V/N_s - I_{total}R_s/N_p)/kT} - 1) - N_pI_{02}(e^{q(V/N_s - I_{total}R_s/N_p)/AkT} - 1) - (N_pV/N_s + I_{total}R_s)/R_{sh} \tag{5}$$

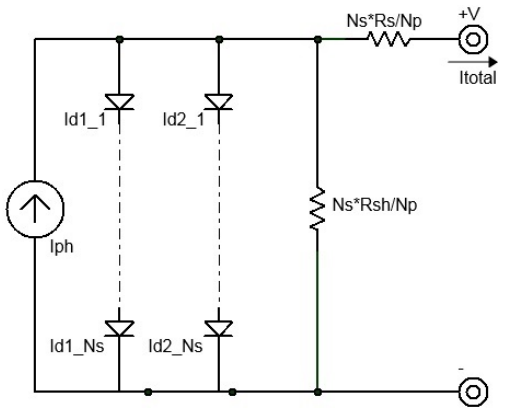


Figure 4: Equivalent circuit for a photovoltaic array.

where I_{ph} is represented by a current source and the current under darkness (I_{01} and I_{02}) by two diodes in parallel, justifying the ideal factor between both diodes denoted as A, N_p and N_s, are the number of cells (from the array) in parallel or series , respectively. The equivalent circuit to (5) is shown in Figure 4.

Using the short-circuit condition, $V = 0$ gives $I_{total} = I_{sc}$, then:

$$I_{sc} = N_pI_{ph} - N_pI_{01}(e_{qI_{sc}R_s/kT}) - N_pI_{02}(e_{qI_{sc}R_s/AkT}) - I_{total}R_s/R_{sh} \tag{6}$$

Meanwhile, using the open-circuit condition, $I_{total} = 0$ gives $V = V_{oc}$, then (5) changes to a:

$$0 = N_pI_{ph} - NpI_{01}(e^{qV_{oc}/N_skT} - 1) - NpI_{02}(e^{qV_{oc}/AN_skT} - 1) - (N_pV_{oc}/N_s)/R_{sh} \tag{7}$$

Combining (6) and (7) gives:

$$NpI_{sc} = I_{01}(e^{qV_{oc}/N_skT} - e^{qI_{sc}R_s/N_pkT}) - I_{02}(e^{qV_{oc}/AN_skT} - e^{qI_{sc}R_s/AN_pkT}) + (V_{oc}I_{sc}R_s)/R_{sh} \tag{8}$$

Differentiating (4) with respect I gives:

$$1 + qN_pI_{01}R_s/kT(e^{q(V/N_s + IR_s/N_p)/kT} + qN_pI_{02}R_s/kT(e^{q(V/N_s + IR_s/N_p)/kT} + R_s/R_{sh} \tag{9}$$

Under the open-circuit condition, $R_{s0} = -dV/dI$, then (9) changes to:

$$(R_{s0} - R_s)(\frac{qN_p}{kT}(I_{01}e^{\frac{qV_{oc}}{N_skT}} + I_{02}e^{\frac{qV_{oc}}{N_sAkT}}) + \frac{1}{R_{sh}}) - 1 = 0 \quad (10)$$

meanwhile, in the short-circuit condition, $R_{sh0} = -dV/dI$, then expression (9) changes to:

$$(R_{sh0} - R_s)(\frac{qN_p}{kT}(I_{01}e^{\frac{qI_{sc}R_s}{N_pkT}} + I_{02}e^{\frac{qI_{sc}R_s}{N_pAkT}}) + \frac{1}{R_{sh}}) = 0 \quad (11)$$

Substituting $I_{total} = I_m$ and $V = V_m$ in (5) and using Iph from (6) gives:

$$I_m(1+\frac{R_s}{R_{sh}}) = I_{01}(e^{\frac{qV_{oc}}{N_skT}} - e^{\frac{q(V_m+I_mR_s)}{N_pkT}}) + I_{02}(e^{\frac{qV_{oc}}{AN_skT}} - e^{\frac{q(V_m+I_mR_s)}{AN_pkT}}) + \frac{(V_{oc}-V_m)}{N_sR_{sh}} \quad (12)$$

Finally, to determine the values of I_{01}, I_{02}, R_s, R_{sh}, A, has been solved the equations system given by (8), (10), (11), (12), with the following conditions that regards the device under test:

$$R_{s0} = 0\ ;\ R_{sh0} = \frac{V_m}{I_{sc} - I_m} - \frac{V_{oc} - V_m}{I_m} \quad (13)$$

3 Experimental Setup

The values of V_{oc}, I_{sc}, V_{max} and I_{max} for the photovoltaic array setup in laboratory, were acquired from I-V curves gives via experimental, under controlled temperature and solar levels, making the necessary measurements and following what is indicated by flow diagram shown in figure 5.

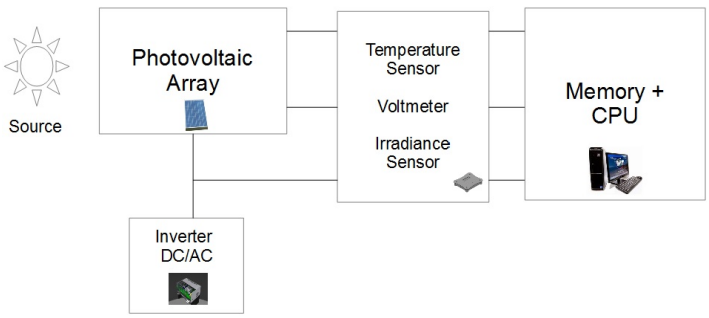


Figure 5: Flow diagram of experimental setup.

In fact, this data was acquired using a device developed in the laboratory based on PIC 18F2550, which has enough memory capacity to operate an SD memory card for data storage, as well as an RS485 interface for trasmission in real time to a computer. The device was based on the scheme proposed in [16]. The device is constituted by three sensors:
- A photodiode to measure the solar irradiance, using the BPW31.
- A temperature sensor, using the LM35.
- A voltmeter, using the ADC PIC module.

To calibrate the solar irradiance sensor, the TES 1333 was used, while for the temperature the infrared thermometer IDEAL 6165 and the multimeter Fluke 179 were used. For the acquisition of data, a script was generated in SCILAB to obtain the data via serial port, in order to plot in real time, for known conditions of irradiance (W/m2) and temperature (^{o} C), I-V curves using a variable load variable (max. 10 A). The figures 6 and 7 show some photos about the experimental setup.

Figure 6: Captures of experimental setup (UNERVEN array).

Figure 7: Captures of experimental setup (SIMENS array).

The technical specifications of photovoltaic (PV) module under test are indicated in table 1: Power (Max.), Voltage (Nominal), Series Cells (Ns) and Parallel Cells (Np).

Table 1: PV arrays under test (both polycrystalline).

Parameter	UNERVEN MSF	SIEMENS PC-2-JF
Power	195 W	35 W
Voltage	24 V	17 V
Ns	48	36
Np	1	1

The I-V curves for several solar irradiance levels during the useful solar day for each model of photovoltaic module under test are shown in figures 8 and 9.

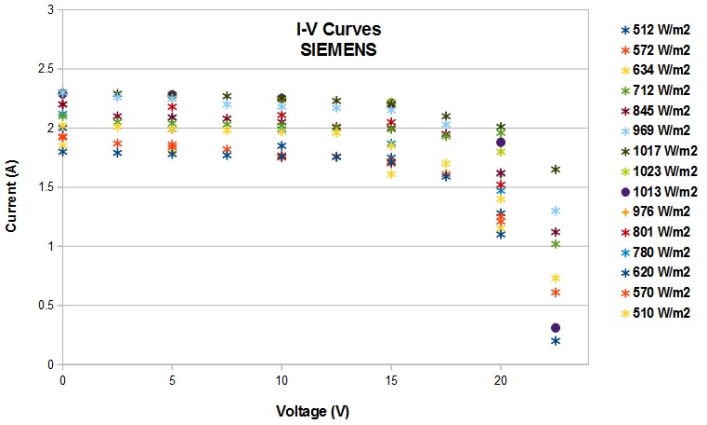


Figure 8: I-V experimental curves for UNERVEN SMF.

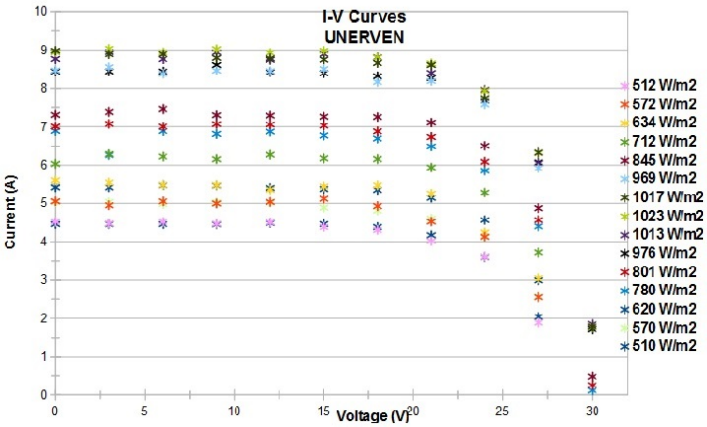


Figure 9: I-V experimental curves for SIEMENS PC-2-JF.

Table 2 shows the values obtained in laboratory for parameters: V_{oc}, I_{sc}, V_{max} and I_{max}, from photovoltaic arrays under test.

Table 2: Experimental values obtained.

Parameter	UNERVEN	SIEMENS
V_{oc}	30.25 V	23.18 V
I_{sc}	8.72 A	2.41 A
V_m	23.76 V	2.12 A
I_m	8.15 A	17.49 V

4 Software Development

A SCILAB script was developed to solve nonlinear equations system constituted by expressions (8), (10), (11), (12), using the function fsolve $[x_0, F]$, with x_0 defined as a vector with the initial values and F defined as a vector that describes equations mentioned before.

First of all, we have to define the equations to solve as a function:

$function[f] = F(x)$
$kB = value;$
$q = value;$
$T = value;$
$Voc = value;$
$Isc = value;$
$Vm = value;$
$Im = value;$
$Vt = value;$
$Rsh = value;$

$f(1) = x(1) * (exp(Voc/Vt) - exp(Isc * x(4)/Vt)) + x(2) * (exp(Voc/(x(3) * Vt)) - exp(Isc * x(4)/(x(3) * Vt))) - Isc + (Voc - Isc * x(4))/Rsh;$
$f(2) = -x(4) * ((x(1)/Vt) * exp(Voc/Vt) + x(2)/(x(3) * Vt) * exp(Voc/(x(3) * Vt)) + 1/Rsh) - 1;$
$f(3) = x(1) * (exp(Voc/Vt) - exp((Vm + Im * x(4))/Vt)) + x(2) * (exp(Voc/(x(3) * Vt)) - exp((Vm + Im * x(4))/(x(3) * Vt))) + (Voc - Vm)/Rsh - Im * (1 + x(4)/Rsh);$
$f(4) = -x(4) * ((x(1)/Vt) * exp(Isc * x(4)/Vt) + (x(2)/(x(3) * Vt)) * exp(Isc * x(4)/(x(3) * Vt)) + 1/Rsh) - 1;$
$endfunction$

Then, we have to define a script to solve the equations:
$x0 = [min_value; scale_factor; max_value];$
$y = fsolve(x0, F);$
$I1 = y(1);$
$I2 = y(2);$
$A = y(3);$
$Rs = y(4);$

The expression defined by (5) is a trascendental equation that was solved in this case using Xcos from Scilab. Figure 10 shows how the four terms in (5) are determined. For more details see [1].

In the last blocks a DC-AC converter has been included and using a full-bridge configuration and a LC filter in the out, with PWM control. For more details in [1].

The I-V and P-V curves generated in Scilab, under Standard Test Conditions (STC) from the mathematical two-diodes model used for each array under test: SIEMENS and UNERVEN, are shown in figures 11 and 12.

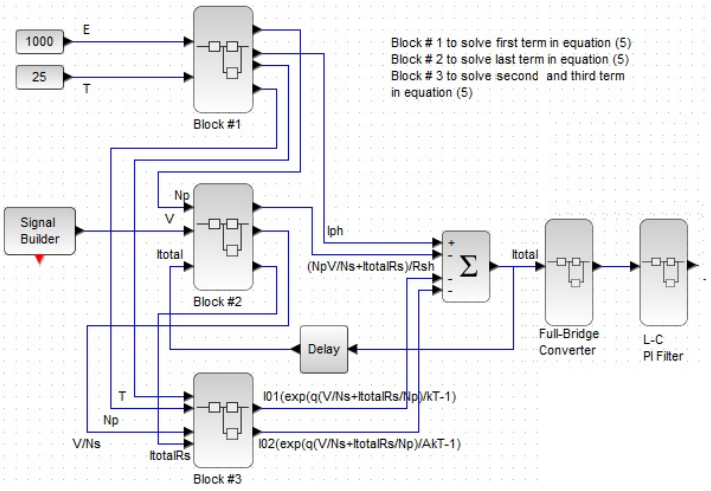


Figure 10: Xcos solution to simulate photovoltaic arrays with DC/AC inverter.

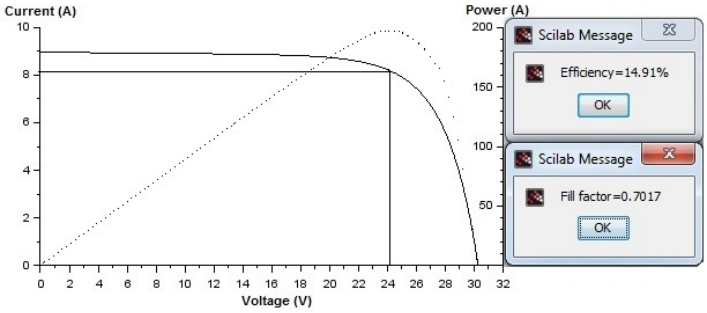


Figure 11: I-V and P-V curves through Scilab (PV array: UNERVEN).

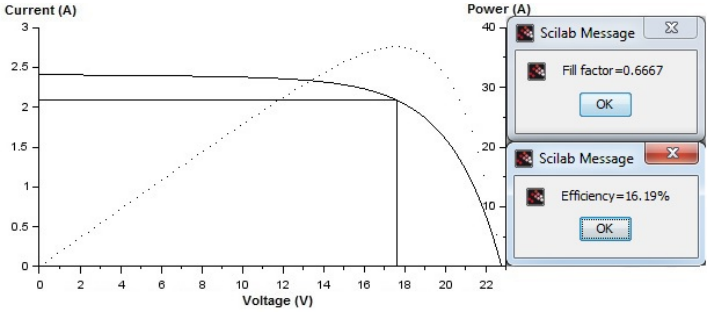


Figure 12: I-V and P-V curves through Scilab (PV array: SIEMENS).

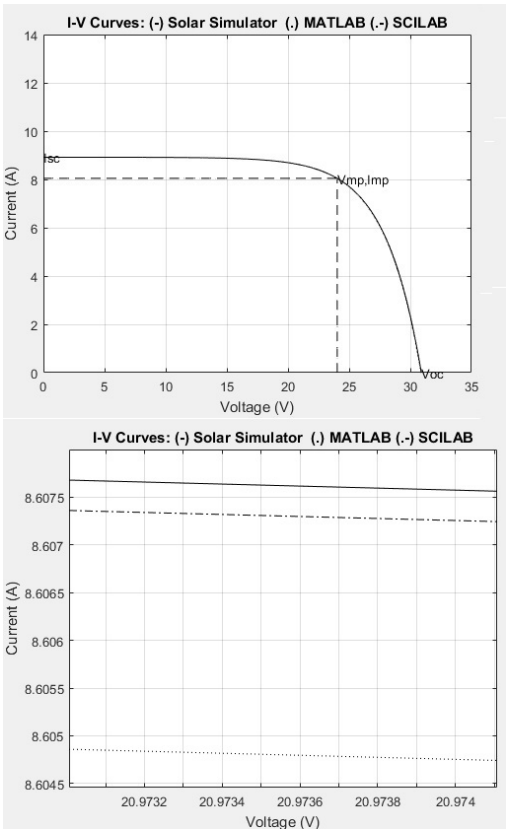


Figure 13: I-V Curves: Solar Simulator, Matlab, Scilab (PV array: UNERVEN).

5 Results

The I-V curves from different sources are shown in Figures 13 and 14: Solar Simulator (from UNERVEN, class A), Matlab Simulator (based on [1]) and Scilab Simulator, for each photovoltaic array under study on STC (1000 W/m^2 , 25 ^oC). The solid line and the dash-dotted line are very close to each other, even if some approximation algorithm was not used to solve the equations of two-diode model as in recent workpapers cited above.

The I-V curves for several solar irradiation levels using Xcos from Scilab for each photovoltaic array under study are shown in Figures 15 and 16. Again, the solid line and the dash-dotted line are very close to each other, with an accuracy superior to 99.9 percent.

6 Conclusions

A computational tool has been put to test in laboratory, based on a free software platform, which allows evaluate in real time the efficiency not only of a photovoltaic array, but also the effects generated by a DC-AC inverter coupled like a load to the system as in real enviroment, using in the develop the physic models cited. Test were coducted in two commercial photovoltaic arrays, and aditionally, the performance results were compared with other tools (licensed) reaching a high accurate.

In contrast with experimental tools used (Solar Simulator Class A from UNERVEN) and computational tool based on Matlab, using Scilab (free software) was improve the perfomance of photovoltaic arrays, even a real time estimation of fill factor and electrical efficiency as an additional feature respect to works cited above. works.

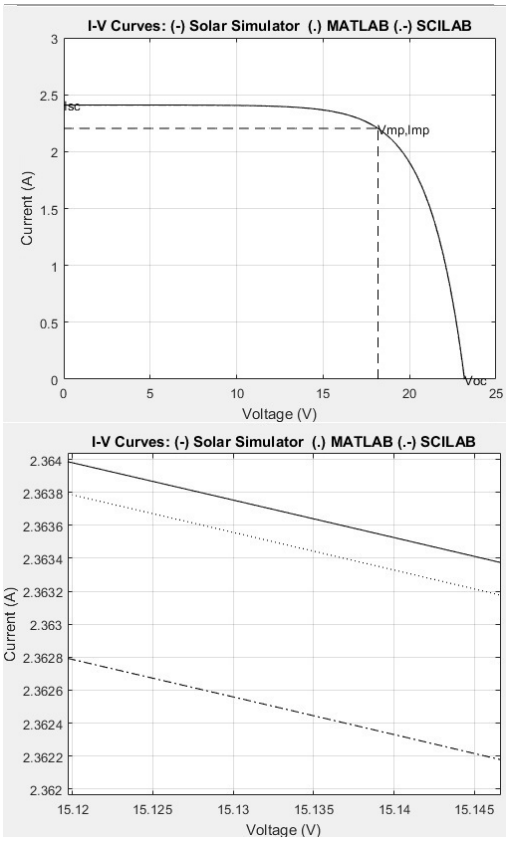


Figure 14: I-V Curves: Solar Simulator, Matlab, Scilab (PV array: SIEMENS).

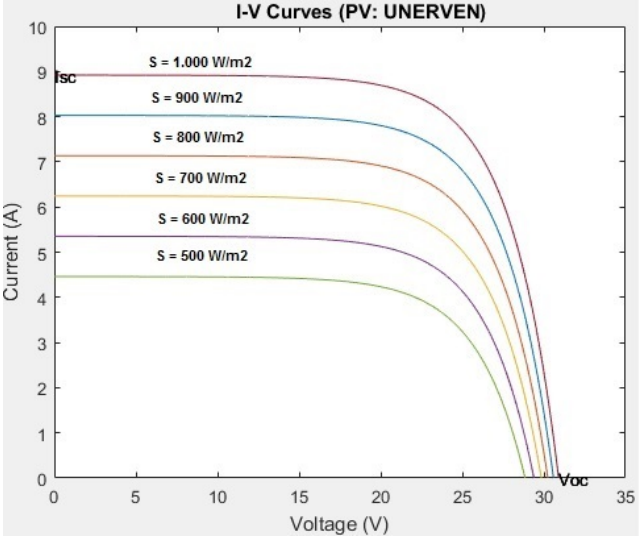


Figure 15: I-V Curves: for several irradiation levels (PV array: UNERVEN).

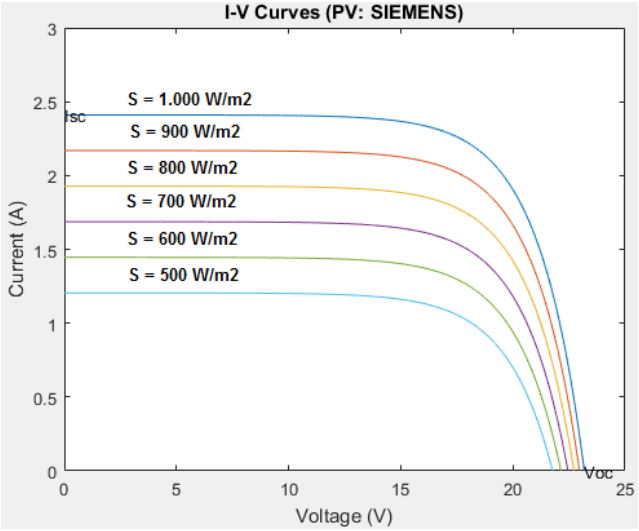


Figure 16: I-V Curves: for several irradiation levels (PV array: SIEMENS).

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment The authors would like to thank to Ministerio del Poder Popular para Ciencia y Tecnología, Ministerio del Poder Popular para la Energía Eléctrica, Unerven - PDVSA Industrial, professors Humberto Rojas and Carlos Rojas from Facultad de Ciencias de la U.C.V.

References

[1] H. L. Núñez Ramírez, G. G. Carvalho Kassar, F. C. Brito Maestre, Y. I. Gamboa Maldonado, L. A. Santos Avendaño, C. Gavidia Toro,"Highly Accurate and Inexpensive Computational Tool for Improving Photovoltaic Arrays Performance and Efficiency Measures" in Proceedings of the IEEE 2018 Open Innovations Conference (OI), Johannesburg, South Africa, 2018. https://doi.org/10.1109/OI.2018.8535899

[2] S. Chowdhury, S. P. Chowdhury, G. A. Taylor, and Y. H. Song, "Modelling, simulation, performance analysis of a PV array in an embedded environment", in Proceedings of the 42nd International Universities Power Engineering Conference (UPEC), , Brighton, UK, 2007. https://doi.org/10.1109/UPEC.2007.4469048

[3] K. Ishaque, Z. Salam, and H. Taheri, "Simple, fast and accurate two-diode model for photovoltaic modules" Solar Energy Materials and Solar Cells, vol. **95**(2), 586-594, 2011. https://doi.org/10.1016/j.solmat.2010.09.023

[4] N. Hashim and Z. Salam,"Critical evaluation of soft computing methods for maximum power point tracking algorithms of photovoltaic systems" IJPEDS,**10**(1), 548-561, 2019. https://doi.org/10.11591/ijpeds.v10.i1.pp548-561

[5] A. Hemza, H. Abdeslam, C. Rachid and N. Aouan,"Simplified methods for evaluating the degradation of photovoltaic module andmodeling considering partial shading" Measurement, **138**, 217-224, 2019. https://doi.org/10.1016/j.measurement.2019.01.098

[6] P. A. Gbadega and A. K. Saha, "Effects and Performance Indicators Evaluation of PV Array Topologies on PV Systems Operation Under Partial Shading Conditions", in Proceeding of the 2019 Southern African Universities Power Engineering Conference, Bloemfontein, South Africa, 2019. https://doi.org/10.1109/RoboMech.2019.8704823

[7] T. Ma, W. Gu, L. Shen and M. Li,"An improved and comprehensive mathematical model for solar photovoltaic modules under real operating conditions" Energy, **184**, 292-304, 2019. https://doi.org/10.1016/j.solener.2019.03.089

[8] P. Ranjan Satpathy, R. Sharma and S. Dash,"An efficient SD-PAR technique for maximum power generation frommodules of partially shaded PV arrays" Solar Energy, **175**, 182-194, 2019. https://doi.org/10.1016/j.energy.2019.03.078

[9] G. Sai Krishna and T. Moger, "Improved SuDoKu reconfiguration technique for total-cross-tied PV array to enhace maximun power under partial shading conditions" Renewable and Sustainable Energy Reviews, **109**, 333-348, 2019. https://doi.org/10.1016/j.rser.2019.04.037

[10] M. Wolf, G. J. Noel, R. J. Stirn, "Investigation of the double exponential in the current-voltage characteristics of silicon solar cells" IEEE Trans. Electron Devices, **24**(4), 419-428, 1977. https://doi.org/10.1109/T-ED.1977.18750

[11] D. S. H. Chan and J. C. H. Phang, "Analytical Methods for the Extraction of Solar Cell Single- and Double-Diode Model Parameters" IEEE Trans. Electron Devices, **34**(2), 286-293, 1987. https://doi.org/10.1109/T-ED.1987.22920

[12] N. Veissid, A. M. De Andrade, "The I-V silicon solar cell characteristic temperature dependence an experimental study using the standard deviation method", in Proceedings of the 10th European Solar Photovoltaic Energy Conference, Lisbon, Portugal, 1991. https://doi.org/10.1007/978-94-011-3622-8_11

[13] A. Gow, C. D. Manning, "Development of a model for photovoltaic arrays suitable for use in simulation studies of solar energy conversion systems", in Proceeding of the 1996 Sixth International Conference on Power Electronics and Variable Speed Drives, Nottingham, UK, 1996. https://doi.org/10.1049/cp:199608909.03.089

[14] Luque, S. Hegedus, Handbook of Photovoltaic Science and Engineering. Willey, 2nd Edition, 2011.

[15] A. Smets, K. Jäger, O. Isabella, R. van Swaaij, M. Zeman, Solar Energyhe Physics and Engineering of Photovoltaic Conversion, Technologies and Systems. UIT Cambridge, 2016.

[16] M. A. Martínez, J. M. Andújar and J.M Enrique, "A New and Inexpensive Pyranometer for the Visible Spectral Range" Sensors, **9**(6), 4615-4634, 2009. https://doi.org/10.3390/s90604615

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 5, 46-52 (2019)
www.astesj.com
Special Issue on Advancement in Engineering and Computer Science

ASTES Journal
ISSN: 2415-6698

Detecting Malicious Assembly using Convolutional, Recurrent Neural Networks

Michael Santacroce[*,1], Daniel Koranek[2], Rashmi Jha[1]

[1]*University of Cincinnati, EECS, 45221, Ohio, USA*

[2]*Air Force Research Laboratory, Sensors Directorate, 45433, Ohio, USA*

ARTICLE INFO

Article history:
Received: 15 March, 2019
Accepted: 7 June, 2019
Online: 06 September, 2019

Keywords:
Malware Detection
Neural Networks
Image Classification

ABSTRACT

We present findings on classifying the class of executable code using convolutional, recurrent neural networks by creating images from only the .text section of executables and dividing them into standard-size windows, using minimal preprocessing. We achieve up to 98.24% testing accuracy on classifying 9 types of malware, and 99.50% testing accuracy on classifying malicious vs. benign code. Then, we find that a recurrent network may not entirely be necessary, opening the door for future neural network architectures.

1 Introduction

This paper is an extension of the work originally presented in the 2018 IEEE National Aerospace and Electronics Conference (NAECON) [1] .

Stakes are always escalating in the arms race that is malware detection. From personal security to breaches at large data centers, it is important to remain vigilant and always explore new methods for combating malware. Classifying executable code images may soon be one of the best methods for malware detection given all of the advances with neural networks in recent times, and we aim to further push the field by questioning the amount of information necessary and size of network necessary to make a judgment.

With our previous work, we proved in concept that using only the .text section of executable files, and not the data, is a possible avenue for malware detection. Using only the executable code allows for leaner, faster networks at the lowest possible level in the hardware stack. Unfortunately, one large issue with malware images is the extremely variable size of the input - code can be anywhere from minuscule to gigabytes in size so figuring out how to use one network to classify code of any size can be difficult. We previously used padding to standardize all files to one length, negatively impacting performance. In this work, we extend on our previous work by using a "windowing" method to allow for arbitrary-length, minimally processed input that achieves competitive performance in malware classification compared with seminal works. To do this, we use a variant of recurrent neural network called a ConvLSTM [2] that uses convolution in tandem with a recurrent architecture. With further investigation, however, we find that it may be possible to ditch the recurrent relationship altogether, opening the door for a multitude of possible architectures.

2 Related Work

Viewing executable code as an image has become a subject of interest for many studying malware detection and neural networks. We will present some of these works for comparison to our work, to both show the state of research and also how our work has some advantages and disadvantages to other methods. In particular, our method has minimal preprocessing of the images, unlike almost all other methods. To view executable code as an image, each byte of the file is taken and translated directly from hexidecimal to decimal. These values then become greyscale pixels, creating an image. Figure 1 shows an example.

Figure 1 shows the first 100 lines of a malicious image where each line of the executable code is translated to a line of greyscale pixels. We pad each line with 0's to the maximum line length. This format differs from some other formats, especially with the line padding and that we only use the .text section, or executable code, and no data. For many of the malware as image works, the results do not classify malicious vs. benign, but instead classifying which type of malware a file is. This is because obtaining and distributing a large number of malicious files is much easier than doing the same

[*]Corresponding Author: Michael Santacroce, santacml on GitHub, santacml@mail.uc.edu

for benign files. Additionally, it is in some ways a better test of the network to classify the type of malware than malicious or benign, as it can be considered a more difficult type of the same problem. If a network can classify the type of a file out of 10 or more classes of executable code, it should be easier to classify out of only 2 classes, in theory.

Figure 1: Example executable code as an image

In 2010, the authors of [3] used k-NN and feature extraction to achieve 98% accuracy on nearly 9,500 samples of 25 classes of malware. This is one of the earliest works that visualize malware as an image and shows that the concept is viable. The authors note some possible methods of avoiding detection such as inserting redundant bytes. Additionally, we note that this method includes a large amount of preprocessing for the malware images.

In [4], the authors process the images to a standard size and then use a densely-connected neural network to achieve 96.35% accuracy on 3131 files of 24 classes of malware. We note that in this work, excellent results are achieved on such a large number of classes of malware.

More than any other method, the ideas of malware image detection have been combined with the burst of developments with convolutional neural networks (CNNs). CNNs have become the go-to for image classification due to their ability to identify spatial relationships in data [5, 6, 7]. In particular, the Microsoft Malware Classification Challenge has spurred a large interest in using CNNs for malware detection [8]. For instance, the author of [9] used a CNN to acheive 95.24% accuracy on a test set of over 40,000 samples, and then used a residual network to achieve 98.21% accuracy on the same samples. The author also notes that the test set was limited to a certain size of binaries and that future work should expand to any length of binaries. In our own work, which this paper extends, we presented up to 88% accuracy in classifying malware on the same data from the Microsoft Malware Classification Challenge [1] . The largest issue from that work was the large amount of padding used. Instead of reducing or expanding images, we padded all images to be the same length. The upside of this method is that it loses no information, however, there were of course large amounts of padding for tiny images, affecting performance. Like in this work, however, we only used the .text section, proving it possible to only need limited information.

The authors of [10] use a convolutional, distributed network, topped with a recurrent neural network on over 2 million malicious files. The network uses bytes of the full executable without preprocessing and uses and embedding, pointing out the issues with preprocessing. The authors also note difficulties with RNNs due to the length of the sequence, as we also found. The network achieves a maximum of 94% accuracy and 98% AUC. This work is the most similar to ours, however, we expand in a few ways. Most importantly, we are only using the .text section of files, a much smaller subset of the total files. Next, we introduce the concept of breaking the malware image into windows, which would theoretically allow for processing only window at a time, saving large amounts of memory. Next, we apply the ConvLSTM architecture to the data which has not been done before, alleviating RNN issues. We also use global average pooling in one of the final network layers to improve generalization [11]. Finally, we are using a different dataset that is much smaller, but we do achieve better overall performance.

The best performance we could find was by the authors of [12], who combine a CNN and LSTM in ensemble, achieving 99.88% accuracy on 40,000 samples balanced equally between malicious and benign files. Then, when used on only the malicious files, they achieved 99.36% accuracy on 9 classes. Again, the input is the full disassembly of the executables and the input is again preprocessed. Additionally, the authors do not present the method of disassembly for the benign files. We will discuss in the next section some possible flaws with mixing these two types of dataset, of which there may be many.

Executable code as an image has become a hot area of research in recent years. In many of these works the images are preprocessed before classification, which might take a large amount of time or resources in practice depending on the operations, and may be bypassable by reverse-engineering the preprocessing. Additionally, all of the works we have seen use the full executable file for classification. By only using the .text section of an executable, classification can be done at a lower level in the hardware stack, with less memory, less time, and less context of what is running. We will address the issues of using arbitrary-length files by using minimal preprocessing and only using the executable code with a ConvLSTM architecture.

3 Network Input

In our previous work we generated our own data using malware obtained from a Windows installation and live malware found online [1] , number 240 total files split evenly between the two classes. There were a few issues with this dataset, namely, the limited size due to difficulty of acquiring samples and the overly-complex tokenization method. The tokenization of assembly resulted in a larger-than-needed vocabulary size and some preprocessing time which we attempting to limit as much as possible.

As a second test, we used the data from the Microsoft Malware Classification Challenge on Kaggle [8], which contains over 20,000 total malware files from 9 classes of malware. With this data, we are able to directly use the machine code and translate it into decimal per byte, which resembles other works as we have discussed. We will again evaluate performance on this dataset. Of these files, 10,869 had labels and then of those 10,375 were found to have .text sections that we could use, however, we quickly hit memory limits while testing longer files. To solve this, we had to limit to only files

that had under 50,000 lines of code, resulting in a total of 8,550 files, of which we used 6,500 training and 2,050 testing files. Because the final goal is classification of malware vs. binary code, however, we have generated an additional dataset consisting of 4,550 benign files from a default Windows installation that were disassembled using IDA pro. Again, due to the 50,000 line cutoff, we were only able to use 3,100 of these files. We then combined the benign data with randomly-chosen malicious files for a second dataset of only malicious vs. benign classification, consisting of 4,000 training and 1,000 testing files.

We will present results on both of these datasets to validate that our own benign files were not biased because of some concerns with the benign data format. While we are comfortable that the datasets are now fair to compare we will show those concerns and how we addressed them. Many benign files were initially identifiable by eye due to some certain characteristics, as in Figure 2:

Figure 2: First 100 lines of a benign code image

The first problem is that there seems to be a long portion of data bytes at the beginning of many benign files that only includes one or two executable bytes. In the Microsoft dataset, these sections are grouped to have many data bytes per line (up to 18). Next, many lines end in the same color pixel. This is because of how we extract data from the disassembled bytecode - once disassembled from IDA, we simply look at every line to see if it starts with ".text", signifying executable code. Then, we extract the hexadecimal bytes from that line until there are no bytes left. Many of these lines end in the byte "db", a signal from IDA that represents the variable size (a byte) and not code. For some reason these extra "db" tokens were never an issue in the Microsoft dataset as there is always some other non-hexadecimal token before them, causing them to be excluded in our extraction. We therefore excluded the "db" at the end of any line (which is always lowercase if it is not an actual byte in the code but added by IDA, so no innocent bytes were lost). With both of these issues it is possible we are disassembling the data differently than Microsoft did in the Kaggle challenge. Because the exact way that the Microsoft dataset was generated is unknown, we present results on both datasets with the assumption that our benign files are not, but may be, biased.

Unfortunately, malware or benignware can come in all sizes, so the question becomes how to approach classifying arbitrary-length files. Additionally, the files can reach extreme lengths, so a straight LSTM or RNN will experience difficulties in training. We therefore split the files into t windows by grouping every w lines of code. The number of windows forms the new number of timesteps, and the dataset can now be seen as an arbitrary-length sequence of images. An example is shown in Figure 3.

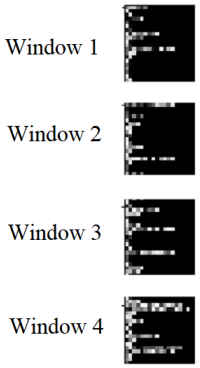


Figure 3: Malware Image Windows

Figure 3 shows a malware image split into windows (not to scale).

4 Network Architecture

Forming the malicious bytecode into an arbitrary-length, windowed format allows for input into a variety of network architectures. Before discussing the network architecture for the new input format, we first downsample and perform an embedding on the data. We previously experienced issues with the size of the input being too large to use at once, and even after purposely omitting files over a certain length, the networks would train in an unreasonably long amount of time. Therefore, we utilized pooling to downsample the input data [1] . While this is the only preprocessing we perform, which adds time to making predictions, we still perform downsampling on the data by randomly discarding 9 of every 10 lines. Along with the benefit of faster training and prediction we find that it does not negatively affect performance and actually improves testing performance in some cases, which we will discuss in the results section.

We also previously used embeddings to assist in feature detection, similar to natural language processing [10, 13], as the magnitude of the bytes have no intrinsic meaning. The embedding serves as a look up table where each singular value maps to a higher-dimensional space of values with less magnitude.

With the initial layer of an embedding, we now expand on the previously static, convolutional network. A natural pick for classification of these videos would be some kind of recurrent neural network given the time-varying nature and seeming temporal dependency of the problem. In fact, one large question we will pose throughout our work is if the windows are actually temporally dependent, in other words - does the previous window affect how we should see the next? Given the nature of the problem, it seems logical that this would be true, and is the assumption we made for beginning to explore architectures.

Convolutional LSTMs are a relatively recent form of recurrent neural network that have cropped up in a few field of study that involve both temporally- and spatially-related data [2, 14, 15]. In

Keras, this layer of a network is called ConvLSTM2d [16], meaning that it performs 2d convolution inside an LSTM. For comparison, A traditional LSTM uses the following equations:

$$\begin{aligned} i &= sigm(h_{t-1} \cdot U_i + x_t \cdot W_i) \\ f &= sigm(h_{t-1} \cdot U_f + x_t \cdot W_f) \\ o &= sigm(h_{t-1} \cdot U_o + x_t \cdot W_o) \\ g &= tanh(h_{t-1} \cdot U_g + x_t \cdot W_g) \\ c_t &= f \cdot c_{t-1} + i \cdot g \\ h_t &= o \cdot tanh(c_t) \end{aligned} \tag{1}$$

Where $W_{i,f,o,g}$ and $U_{i,f,o,g}$ are trainable weight vectors, and c_t and h_t are the hidden state and output, respectively. These equations can of course change and are only an example for comparison.

LSTMs can help alleviate "vanishing gradient" issues when training and therefore allow for training recurrent networks on much longer inputs than a vanilla RNN [17, 18]. We attempted to use a regular LSTM on input mawlware without windows and found that the model would not only take an extraordinarily long time to train, as expected due to the input length, but would not perform well. Even if a decent training accuracy was obtained the results would not translate to testing accuracy.

Recent works, including our own, have shown that classifying malware images responds well to convolution [1] . Therefore, instead of using a LSTM, it is possible to use the Convolutional LSTM (from now on, ConvLSTM). The ConvLSTM consists of nearly the same parts as a regular LSTM except that many dot products are simply switched to being convolution:

$$\begin{aligned} i &= sigm(h_{t-1} * U_i + x_t * W_i) \\ f &= sigm(h_{t-1} * U_f + x_t * W_f) \\ o &= sigm(h_{t-1} * U_o + x_t * W_o) \\ g &= tanh(h_{t-1} * U_g + x_t * W_g) \\ c_t &= f \cdot c_{t-1} + i \cdot g \\ h_t &= o \cdot tanh(c_t) \end{aligned} \tag{2}$$

Where $W_{i,f,o,g}$ and $U_{i,f,o,g}$ are trainable weight vectors, and c_t and h_t are the hidden state and output, respectively.

The ConvLSTM was introduced in 2015 by [2] as a way to capture both spatial and with temporal relations in precipitation forecasting, as opposed to an LSTM, which is only capable of capturing temporal relations. ConvLSTMs have since been used many times to capture spatio-temporal relations in data. In [15], they are used to build on feature extraction in gesture recognition, and in [19], the authors add to the ConvLSTM architecture for traffic accident prediction. More notably, in [20], the authors use a nested ConvLSTM for video classification. Our work can be seen as converting the malware to a video rather than an image, and taking inspiration from these works. In our case, the ConvLSTM captures spatial relations in blocks and functions of code and temporal relations between those blocks.

With the recurrent networks it was found best to use use two hidden recurrent layers, one that returns sequences and a second that does not. When a recurrent network in Keras has the parameter return_sequences set to true, the output of the layer is not just the final output but the output of each timestep, the result being a sequence of the same length. Using this parameter is useful as it allows the stacking of recurrent layers and therefore deeper networks.

While our models that use ConvLSTM found success, which we will discuss in the results section, we also found that they took up a large amount of time to train due to the complexity of the hidden state. Many modified RNNs were attempted by removing or adding various gates to the ConvLSTM hidden state until we tested a recurrent network that simply adds the previous hidden state to the current as so:

$$\begin{aligned} h_t &= x_t * W_t + b_t + h_{t-1} \\ y_t &= \sigma(h_t) \end{aligned} \tag{3}$$

Where σ represents the sigmoid function, W_i and b_i are trainable weight vectors, h_t is the hidden state, and y_t is the output at time t.

This network is about as minimal as a recurrent neural network can get and can be compared to a standard recurrent network where the hidden weight vector is set to all 1s. We call this the "MinConvRNN". In theory, one layer of this network that does acts the same as if each timestep was independently run through a small neural network and then summed afterwards. Because there are two hidden layers used and each timestep is returned for the first, however, the intermediate timesteps returned by the first hidden layer do not resemble a kind of distributed network.

We can perform backpropogation on the MinConvRNN for training through the widely-known procedure of BackPropogation Through Time (BPTT). We will perform BPTT on the MinConvRNN for completeness, however, gradients are calculated symbolically and automatically in our simulations, so these calculations are not stricly necessary. BPTT consists of unrolling the recurrent neuron into t neurons and repeatedly applying the chain rule. We will start by defining all parameters in Table 1, for any arbitrary objective function J.

Table 1: MinConvRNN Backpropogation Parameters

Variable Name	Purpose
x_t	Input at time t
h_t	Hidden state at time t
W_t	Weight vector at time t
b_t	Bias vector at time t
y_t	Output of neuron at time t
σ	Sigmoid Function
δ_t	$\frac{\partial J_t}{\partial y_t}$
J_t	Objective function value at time t

The goal of backpropogation is to find the partial derivative of the objective function with respect to a parameter. In this instance, $\frac{\partial J_t}{\partial W_t}$. Note that W_t for time t is a formality, and $W_i = W_j, \forall (i, j) \in t$, and the same for b_i. We begin by defining the base case of $t = 1$ such that $h_0 = 0$, making $h_1 = x_1 * W_1 + b_1$.

Then, we can define the partial derivative of J with respect to W_t using the chain rule:

$$\frac{\partial J_t}{\partial W_t} = \frac{\partial J_t}{\partial y_t} \cdot \frac{\partial y_t}{\partial h_t} \cdot \frac{\partial h_t}{\partial W_t} \tag{4}$$

We can define δ_t to be $\frac{\partial J_t}{\partial y_t}$ to simplify notation as the expression will change for a different J. We can then calculate the other terms directly as:

$$\begin{aligned} \frac{\partial y_t}{\partial h_t} &= \sigma'(h_t) \cdot 1 \\ \frac{\partial h_t}{\partial W_t} &= \frac{\partial x_t * W_t}{\partial W_t} \end{aligned} \tag{5}$$

We leave $\frac{\partial x_t * W_t}{\partial W_t}$ alone to simplify the expression, but we note it is computable. It is also known commonly known that $\sigma'(x) = \frac{\partial \sigma(x)}{\partial x} = \sigma(x) \cdot (1 - \sigma(x))$, and we will use $\sigma'(x)$ for simplification of expressions as well.

The full partial derivative at time t can then be defined as:

$$\frac{\partial J_t}{\partial W_t} = \delta_t \cdot \sigma'(x_t * W_t + b_t + h_{t-1}) \cdot \frac{\partial x_t * W_t}{\partial W_t} \tag{6}$$

Because W_t is the same $\forall t$ as stated earlier, we can compute the partial derivative for W for n timesteps as:

$$\begin{aligned} \frac{\partial J}{\partial W} &= \sum_t \delta_t \cdot \sigma'(x_t * W + b_t + h_{t-1}) \cdot \frac{\partial x_t * W}{\partial W} \\ h_0 &= 0 \\ t &\in [1, 2...n] \end{aligned} \tag{7}$$

For a recurrent neural network where only the output is considered, we could factor out δ_t from the sum. We can also perform a similar process to find the partial derivative for b for n timesteps as $\frac{\partial J}{\partial b} = \sum_t \delta_t \cdot \sigma'(x_t * W + b_t + h_{t-1})$, with the same conditions.

This partial derivative does not look too dissimilar to that of a basic recurrent neuron. The neuron is indeed recurrent in the sense that it is necessary to calculate h_{t-1} in order to calculate h_t, and that the gradient must be propogated starting with the initial output in the same manner. Regardless of these equations, however, the neuron as defined in Eq. 3 does not make intuitive sense. Taking the raw sum after convolution, in theory, might simply lead to an unbounded output in the hidden state that approaches ∞ (or $-\infty$) for all elements as the number of timesteps approaches ∞ , leading to all 1s (or 0s) in the output with the sigmoid activation.

As will be shown in the results, we found the MinConvRNN to largely match the performance of the ConvLSTM2d. The success of the MinConvRNN begs the question if a recurrent neural network is necessary, or in other words, if executable code as an image has a temporal dependency when it comes to classification. As will be presented in results, it is even possible to scramble the order of the input lines and achieve decent performance using the same net trained on ordered input lines. It seems unintuitive at first but a simple explanation is simply that a large amount of code is wrapped up in short functions. Depending on how large the window size is made, it is possible that each window encapsulates even multiple functions. Because the order in which functions are defined largely does not matter, there is not a strong temporal dependence between windows. A different explanation may be that even between windows, temporal dependency does not matter, and that the class of a file is largely determined by the types and quantities of commands issued, and not their order. Future work will need to investigate the temporal dependence of the data more closely as it will greatly affect how we classify malware with neural networks.

Regardless of the type, after the recurrent layers, the output has three dimensions. The first two dimensions can be average-pooled to reduce overfitting [11], and then connected with a final densely-connected layer for the output. The architecture of the network is shown in Figure 4.

5 Performance

We present results on both the Malware Classification Challenge dataset and the Malware vs. Benign dataset as discussed, starting with the Malware Classification Challenge dataset.

Table 2 shows the final results of the ConvLSTM2d and MinConvRNN networks. Both networks were trained until no further performance gains were seen.

Table 2: Results on Microsoft Malware Classification Challenge Dataset

Metric	ConvLSTM	MinConvRNN
Parameter Count	573,634	116,951
Train Accuracy	98.35%	97.28%
Test Accuracy	98.24%	96.39%
Time to Inference	~63.67ms	~23.73ms

In Table 2, it can be seen firstly that both networks performed competitively to other works. The MinConvRNN was purposely limited in size to have the same number of convolutional kernels as the ConvLSTM2d network, being 50 kernels of size 4 in the first hidden recurrent layer, and 25 kernels of size 3 in the second hidden recurrent layer. The network was kept the same size to demonstrate the relatively similar performance. While the ConvLSTM2d network did achieve better results, it has nearly 5 times the total number of parameters. Better results are certainly possible through tweaking the network size, activation functions, etc., but we see these networks as a proof-of-concept for future work that does not rely on recurrent networks given the performance of the MinConvRNN. Additionally, we note that the ConvLSTM2d takes nearly twice as long to make a prediction given an input due to the complexity of the hidden state. If we can cut out the recurrent nature of the network, the time to predict would drastically go down.

Table 3 shows Confusion matrix for the ConvLSTM2d network on the testing data, where the row is the actual class and the column is the predicted class. Of note is that one class in the dataset has an extremely low number of samples, class 5. Neither this network nor the MinConvRNN was able to correctly identify and class 5 samples. This might be addressed by using weighted categorical crossentropy, but again, we are presenting these results as a proof-of-concept.

Table 4 shows the confusion matrix for the MinConvRNN on the testing dataset, where the rows are the actual class and the columns are the predicted class.

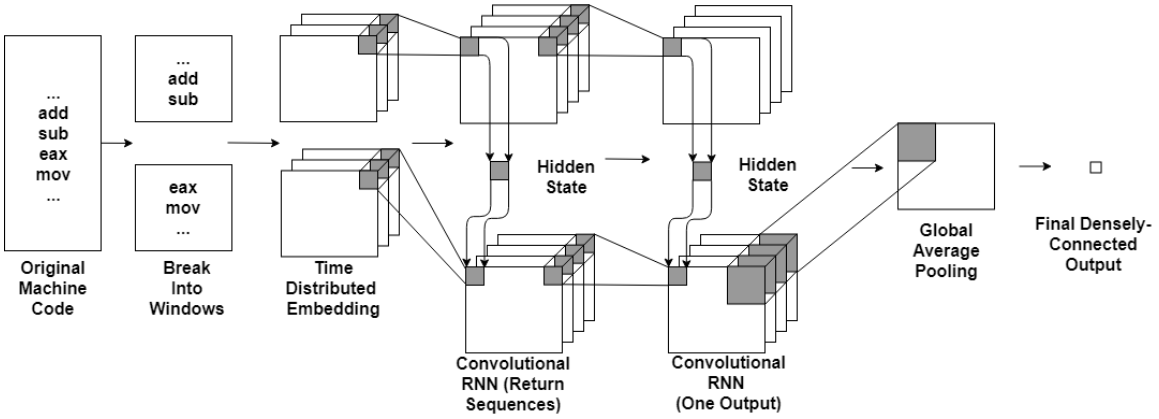


Figure 4: RNN for Malware Classification Architecture (for arbitrary hidden state)

Table 3: Confusion Matrix for ConvLSTM2d Net on Testing Data

	1	2	3	4	5	6	7	8	9
1	249	0	0	0	0	0	0	0	0
2	2	364	0	0	0	0	0	0	0
3	0	0	681	0	0	0	0	0	0
4	0	0	0	99	0	0	0	0	0
5	2	1	0	0	0	0	1	2	0
6	3	0	0	3	0	107	0	0	0
7	0	1	0	0	0	1	89	2	1
8	6	1	0	0	0	1	0	259	1
9	3	1	0	0	0	0	0	4	166

Table 4: Confusion Matrix for MinConvRNN on Testing Data

	1	2	3	4	5	6	7	8	9
1	245	0	0	0	0	0	0	2	2
2	3	356	1	0	0	2	0	3	1
3	0	0	680	0	0	0	0	1	0
4	0	0	1	98	0	0	0	0	0
5	1	0	0	1	0	1	0	2	1
6	1	0	0	4	0	105	0	2	1
7	0	1	12	0	0	1	80	0	0
8	13	2	0	1	0	0	0	252	0
9	4	1	5	1	0	1	0	2	160

Overall, we see performance for both networks, moreso the ConvLSTM2d, matching other works. The benefit to using these networks is the minimal preprocessing, the small size of the networks, and the minimal amount of input by only using the executable code.

Table 5 shows the results of the networks on the Malicious vs. Benign dataset.

Results on the Malicious vs. Benign Dataset are on par or better than the previous dataset. Despite reservations with the benign data as previously discussed we believe that the networks show competitive performance in classifying Malicious vs. Benign data to seminal works. The MinConvRNN again shows worse performance than the ConvLSTM, likely due to the smaller number of parameters. Considering that the network is smaller, the relative success of the MinConvRNN calls into question whether using a recurrent network is necessary.

Table 5: Results on Malicious vs. Benign Dataset

Metric	ConvLSTM	MinConvRNN
Parameter Count	573,634	116,951
Train Accuracy	99.78%	99.32%
Test Accuracy	99.50%	99.20%
Test Recall	99.37%	98.12%
Test Precision	99.58%	99.58%
Time to Predict	~91.16ms	~35.32ms

6 Future work

There are many topics of future work to expand on. In the big picture, the concept of a non-recurrent network would be interesting to test compared to the recurrent networks. It is possible to take more inspiration from video classification methods in this way. We have observed that a minimal recurrent network achieves decent performance, so with optimization, a purely time-distributed network that is aggregated afterwards may perform well.

With or without recurrent, it would be important to optimize the network, exploring activation function, weighted categorical crossentropy, cross-validation methods, and pruning methods. Performing an architecture optimization search would take much more work, but would improve the performance of the network.

There are also some flaws in our methodology, the largest being the limit of 50,000 line long samples. One possible method of removing this constraint is to reformat the data. The figures of malware code show that each line is padded with 0's to be the same length - this was done to preserve the ordering of the opcodes in relation to the data. This format, however, may be unnecessary and the code is often treated as one contiguous block in other works. The size of the input would be greatly reduced by removing the padding, which would not only allow for any length input, but also faster and possibly more accurate predictions. Additionally, it would also

ensure compatibility with the Microsoft and Benign datasets given the different maximum line lengths.

Finally, finding a way to extend the work to hardware would be great for evaluating the feasibility of the practical implementation of the networks. One possible method of using the malware-detecting-network in practice might be to deploy it in a custom chip or FPGA. While the network is likely too large for such acceleration now, with compression techniques, it may be possible to make the network fit. Putting the network in separate, purpose-built hardware would not only free up resources from the main system but also speed up the network considerable.

7 Conclusion

Malware detection will always be a vital part of computer systems. In recent times, classifying malware images made from the executable bytes of a file has become a possible route to take.

This extension of our previous conference paper again proves it is possible to classify malware with nothing other than the .text section of an executable, which contains the actual code that makes up the program. Additionally, we expand on the type of network for detection by creating windows out of the malware image. With this format, minimal preprocessing is required to feed the image into a convolutional, recurrent neural network, achieving competitive performance.

Additionally, we find that using a recurrent neural network may not be entirely necessary. By stripping the ConvLSTM2d of its components and making the MinConvRNN, we have shown that the temporal dependency between windows may not be necessary, opening the doors to new network architectures. Using the MinConvRNN is not the end of the story - it is highly likely that even leaner neural network architectures are possible for malicious assembly.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment The authors would like to thank the University of Cincinnati, Dr. David Kapp at AFRL, WPAFB, OH, and AFRL GRANT No. 1014236 under subaward No. 1919.05.05.91 with TDKC.

References

[1] M. Santacroce, D. Koranek, D. Kapp, A. Ralescu, and R. Jha, "Detecting Malicious Assembly with Deep Learning," in *NAECON 2018 - IEEE National Aerospace and Electronics Conference*. IEEE, jul 2018, pp. 82–85. [Online]. Available: https://ieeexplore.ieee.org/document/8556657/

[2] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, W.-C. Woo, and H. Kong Observatory, "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting," Tech. Rep. [Online]. Available: https://arxiv.org/pdf/1506.04214v1.pdf

[3] L. Nataraj, S. Karthikeyan, G. Jacob, and B. S. Manjunath, "Malware Images: Visualization and Automatic Classification," Tech. Rep., 2010. [Online]. Available: https://vision.ece.ucsb.edu/sites/vision.ece.ucsb.edu/files/publications/nataraj{_}vizsec{_}2011{_}paper.pdf

[4] A. Makandar and A. Patrot, "Malware analysis and classification using Artificial Neural Network," in *2015 International Conference on Trends in Automation, Communications and Computing Technology (I-TACT-15)*. IEEE, dec 2015, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/document/7492653/

[5] W. Rawat and Z. Wang, "Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review," *Neural Computation*, vol. 29, no. 9, pp. 2352–2449, sep 2017. [Online]. Available: http://www.mitpressjournals.org/doi/abs/10.1162/neco{_}a{_}00990

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," pp. 1097–1105, 2012. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks

[7] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale Video Classification with Convolutional Neural Networks," Tech. Rep. [Online]. Available: http://cs.stanford.edu/people/karpathy/deepvideo

[8] R. Ronen, M. Radu, C. Feuerstein, E. Yom-Tov, M. Ahmadi, and M. . Crowdstrike, "Microsoft Malware Classification Challenge," 2018. [Online]. Available: http://doi.acm.org/10.1145/2857705.2857713,

[9] A. Singh, "Malware Classification using Image Representation," Tech. Rep., 2017. [Online]. Available: https://pdfs.semanticscholar.org/5bba/bd377616bd446a194f432a12e98326f5c637.pdf

[10] E. Raff, J. Barker, J. Sylvester, R. Brandon, B. Catanzaro, and C. Nicholas, "Malware Detection by Eating a Whole EXE," oct 2017. [Online]. Available: http://arxiv.org/abs/1710.09435

[11] M. Lin, Q. Chen, and S. Yan, "Network In Network," dec 2013. [Online]. Available: https://arxiv.org/abs/1312.4400

[12] J. Yan, Y. Qi, and Q. Rao, "Detecting Malware with an Ensemble Method Based on Deep Neural Network," *Security and Communication Networks*, vol. 2018, pp. 1–16, mar 2018. [Online]. Available: https://www.hindawi.com/journals/scn/2018/7247095/

[13] T. Schnabel, I. Labutov, D. Mimno, and T. Joachims, "Evaluation methods for unsupervised word embeddings," Tech. Rep., 2015. [Online]. Available: http://www.cs.

[14] J. Donahue, L. A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, and T. Darrell, "Long-Term Recurrent Convolutional Networks for Visual Recognition and Description," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 677–691, apr 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7558228/

[15] G. Zhu, L. Zhang, P. Shen, and J. Song, "Multimodal Gesture Recognition Using 3-D Convolution and Convolutional LSTM," *IEEE Access*, vol. 5, pp. 4517–4524, 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7880648/

[16] F. Chollet, "Keras Documentation." [Online]. Available: https://keras.io/

[17] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, nov 1997. [Online]. Available: http://www.mitpressjournals.org/doi/10.1162/neco.1997.9.8.1735

[18] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A Search Space Odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, oct 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7508408/

[19] Z. Yuan, X. Zhou, and T. Yang, "Hetero-ConvLSTM," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '18*. New York, New York, USA: ACM Press, 2018, pp. 984–992. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3219819.3219922

[20] C. Lu, M. Hirsch, and B. Schölkopf, "Flexible Spatio-Temporal Networks for Video Prediction," Tech. Rep. [Online]. Available: http://openaccess.thecvf.com/content{_}cvpr{_}2017/papers/Lu{_}Flexible{_}Spatio-Temporal{_}Networks{_}CVPR{_}2017{_}paper.pdf

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 53-60 (2019)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Proposal for a Security Model for a Popular Voting System Process in Latin America

Segundo Moisés Toapanta Toapanta [*,1], Allan Fabricio German Diaz[1], Darío Fernando Huilcapi Subia[1], Luis Enrique Mafla Gallegos[2]

[1]*Department of Computer Science, Salesian Polytechnic University (UPS), Guayaquil, Ecuador*

[2]*Faculty of Systems Engineering, National Polytechnic School of Ecuador (EPN), Quito, Ecuador*

ARTICLE INFO

Article history:
Received: 08 July, 2019
Accepted: 21 August, 2019
Online: 03 September, 2019

Keywords:
Technology
Electronic Vote
Electronic Security
Technological systems
Individual reinsurance

ABSTRACT

In Latin America, there is a need for better security systems in different national or sectional votes, which is why this study was carried out, to show a proposal for a security model that helps to obtain better results and to this purpose has been selected information on the database for different Latin American countries, it has also been considered descriptive aspects and with a quantitative approach, to provide an objective livelihood. It has been intended to identify several security requirements that give confidence to citizens. Currently in Latin America studies continue to be carried out on technological systems for electronic voting and their real effectiveness in different countries, and then improve or implement them. Not all have come to implement it completely but in part due precisely to the demands it demands, also because of the limited information and because there are few related studies on these issues. The growth of technology and its easy access has allowed it to be used in different areas and politics is no exception, which is why different modalities of electronic voting have been implemented so that they can reduce some failures that are maintained in the traditional voting system. In conclusion, it is noted that voters need to know better about the new systems applicable to counting, it encourages individual reinsurance, which does not reveal the identity of the voter, as well as transparency, easy access information and auditability. Awareness is suggested about the use of electronic voting and the benefits it brings without also excluding negative factors because there is not totally secure and reliable digital voting system. Citizen participation must go hand in hand with political decision-making, so that processes are transparent, and results are protected. With all these criteria it can be noted that it is necessary to consider a comprehensive electronic security model that guarantees that reliability and integrity of the secret election vote, bearing in mind that the purpose is to reduce and prevent errors.

1. Introduction

Several Latin American countries are implementing new technologies and wish to implement it in the electoral system through electronic voting. Few countries are already using it with relative success, as each electoral process receives serious criticism, even in Venezuela the results achieved in the last elections have been rejected. These events require a thorough analysis, to know if the process has been inconvenient or the problem has been in pre-election events, which have nothing to do with existing technology platforms.

Author Diaz Ricardo refers that the momentous changes in the modern world, characterized by its interesting development, the accelerated globalization of the economy, the sharp dependence that incorporates a high volume of information and the systems that provide it; the increase in vulnerability and the broad spectrum of threats impose new challenges on the practice of the audit profession, in particular to the computer security audit [1].

The use of ICTs in Latin America has been increasing and it has been desired to be applied in electronic voting to generate an effective and transparent information process. E-voting or electronic network voting is an applicable mode via the Internet, as well as the electronic ballot box that has the most acceptance and use in Latin America compared to the other existing modalities.

* Segundo Moisés Toapanta Toapanta, Email: stoapanta@ups.edu.ec

If we are thinking of implementing a security system, it is necessary to establish the voting mechanism that best meets the requirements of each country, as well as to think about the scopes and budgets that are needed for its applicability. The author Muñoz points out that organizations need a stability and greater degree of protection focused on computer security to protect and minimize threats to their information [2].

Considering the benefits and potential expenses make the decision more difficult to make, as well as having people who handle these systems with the greatest care that these processes deserve, so the challenge of using technological aspects requires that citizens and the authorities in charge maintain ethical criteria to avoid the different types of fraud that may change the results obtained. Process management and the application of security systems will make security systems applicable to each Latin American country, under any type of government.

2. Materials and Methods

2.1. Materials

The material used has been selected based on data relating to different Latin American countries, among these, articles based on the implementation of different modalities of electronic system and the benefits it provides. Current work was also considered to make security models applied to electoral systems relevant.

The information has been relevant and closely related to the subject matter of this study, to consider the urgency of presenting a security system in the voting system of Latin American countries. Several Latin American countries have seen suffrage security measures applied in other countries in the region, while the will to implement electronic voting is being considered, including Argentina, Chile, Peru, Uruguay and Costa Rica.

In the region only Brazil and Venezuela show the use of electronic voting, but that as in any electoral system, there are reports of the results obtained. According to Hosp and Vora's theorem, there is no electoral system with perfect integrity, verifiability and privacy, as malicious attempts to modify the results will always be found.

But you can count on important suggestions such as individual reinsurance, which does not reveal the identity of the voter, as well as transparency, with easily accessible information and auditability, that allows access to the result obtained, appropriate use of technological equipment that does not keep a voter's record, the key safeguarding, which in the case of using cryptography, must specify how and who will be responsible and also restrict access to the different equipment used.

2.2. Methods

This work has a quantitative approach since information is presented in an objective, graphical way and making use of available resources. It also considers inductive, deductive and correctional aspects, with descriptive criteria of the process to be followed in the proposed safety system.

For this reason, the studies carried out in this topic are valued, in relation to technological resources, since they are being used in different areas that demand their immediate application.

I) Electoral Information and Technological Modernization

Technological innovation has a great impact and acceptance worldwide especially in Latin America. This implementation in the different countries has provided several benefits in the different processes that make use of technological means.

The author Curious to make most of the benefits of technology, the State requires an appropriate physical infrastructure that supports it through a large transport and access network that allows it to integrate into other networks [3].

So also, the author Jerez points out that the contemporary discussion on public spheres has been expanding the focus of analysis focused on social and political actors towards that of institutional and cultural regulations in public spaces and as key areas of production of ideologies, identities and agendas. To a large extent on these discursive plots, the social conflict and institutional dialogue that processes collective problems is deployed, demeaning which will be defined as public and private interests to be addressed (or not) by each State in a context national history given [4].

In the public and private spheres, they require the development of appropriate technologies, because of the easy access to the information they provide and in addition to their easy management, taking into account that ICTs have also reached each person through the mobile phones, the internet, among others, in such a way that their use is daily and the connection with the different processes of each organization keeps them close.

For his part the author Rovelli, several countries in Latin America were the scene in the last 15 years of intense socio-political and techno-economic changes driven from different policy trends, which converged in principle and to the less from a normative level in a greater presence of the State in the social sphere and the search for alternative and innovative ways for development [5].

Technological innovation, as well as taking part in different organizations in which they provide services such as health, education, entertainment, etc., has become part of the political processes of each country, as in the voting system via Electronic. This has benefits, but there are also risks. An example of electronic security problems occurred in the USA, so author Ottoboni refers that the State of Georgia was a focal point in the civil rights movement of the twentieth century. also has a history of electoral problems: suppression systematic voter voting machines that are vulnerable to undetectable security breaches, and serious security breaches of their data systems [6].

Likewise, in addition to the agility provided by electronic voting, transparency is needed throughout the process, as there are factors that can have a direct or indirect impact during the process or on the use of information.

That is why it is essential to have a security model that is applied in each electoral process in such a way that it can be replicated in other Latin American countries. These technological means also seek to prevent any act of corruption that alters results in such a way that the level of transparency in electoral processes is evident to all voters.

The use of ICTs in electronic voting is intended to generate an effective and transparent information process, which is why there are currently many modalities that are in line with this. A good development of the electoral process through the automation of the vote will also depend on the security model that it uses in order to achieve a correct and adequate management of the information, whatever the mode of the vote to be used.

The different governments of Latin America have decided to implement the automation of the electoral process in order to reduce time, space, resources, among other aspects; considering also the benefits it provides and the great changes it has brought through its use in such a way that citizen participation increases and the voting system is easy and accessible to all.

Like any electronic voting system, they must rely, in dispensing with, security requirements that give voters confidence, make good use of stored information, take care of personal data, and keep the right to integrity of the voter and their decision on the vote.

Studies on electronic voting systems and their real effectiveness in different countries are still being carried out. Not everyone has come to implement it completely but partially because of the lack of security it has in terms of it because of its limited information and related studies on them.

Little information that needs to be reduced in society in such a way as to raise awareness of the use of electronic voting and the benefits it brings without also excluding negative factors because there is no totally secure digital voting system and reliable.

II) Electronic Voting Modalities

Thanks to the technological modernization to which the world has become, mechanisms applied to the electoral field have been observed, with multiple systems that provide what is expected in an electronic voting system: Trust, security, among other characteristics, that a system must rely on, in addition to reducing the costs of the required election materials.

Author Mellinghoff believes that when voting there must be the principle that elections should be secret, it ensures that the voter is the only one who knows the content of his election decision.

This principle is the most important industrial protection for freedom of choice. It obliges the legislator to take the necessary measures to protect the secret ballot (e.g. covered ballot, booths that guarantee voter privacy).

Any investigation of voting intentions or voting of voters is prohibited. Any exception to the secret ballot is permitted only if it favours the right to vote, that is, in cases where the principle of general and equal elections requires a procedure in which it cannot be guaranteed in each case that the vote is absolutely secret [7].

However, it is considered that these are not exempt from any error or failure situation due to some factors to which they may be vulnerable and become a threat thus causing the alteration and loss of information.

According to the author Schmidt, it refers that, in its broadest form, electronic voting refers to the use of electronic technology for the automation of an electoral process [8]. But not just any technological system, but that which applies to the Ecuadorian context.

The author Sandoval indicates that theoretically electronic voting on the internet can be considered a more form of e-vote, so it can be considered that there were some precedents in this regard [9].

There are two types of electronic voting:

- Remote electronic voting.

  Emission of votes through any device (computer, mobile phone, PDA, etc.) with internet connection.

- Electronic voting in person.

  Issuance of Votes from electronic voting terminals from the polling stations (JRV).

It is necessary to emphasize that there are several models of electronic voting in person, among which the electronic ballot boxes stand out, these are carried out in person, that is, located in each electoral roll.

The e – electronic voting or electronic network voting is done via the internet; the reading is automated and can be seen in Figure 1 below.

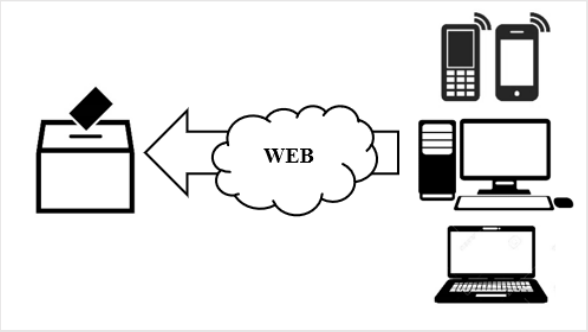


Figure 1: Electronic Vote Models

The electronic urn has the greatest acceptance and use in Latin America compared to other existing modalities; it is carried out in person where there are authorities responsible for the surveillance and control of the votes, in the ballot box for their proper management and organization.

On the other hand, when using digital instruments or technological equipment to represent the use of ballot boxes, it is considered the introduction of elements that are similar in their application but that allow the management of information in a safe, transparent, fast and practical.

For the vote to be served, electronic ballot boxes and the internet are required at the polling place, as shown in Figure 2.

These machines automatically identify the votes where the count is made after this and then be sent to a center that collects all the votes obtained from the polls throughout the country or city and finally obtain the final result of the electoral process.

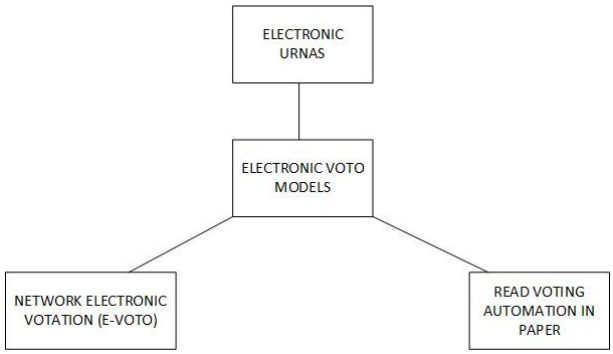


Figure 2: Main Mode Used

Author Abdala indicates that empirical evidence on the extent to which different voting technologies affect election results is far from conclusive. However, irrespective of the causal mechanism used to explain the results, there is a kind of consensus on the importance of the voting system. It is necessary to understand how voting technologies interact with citizens to model political behavior [10].

Electronic voting systems have their advantages and benefits aimed at an entire society not only because of its easy and convenient access or handling, but also because this system avoids long hours of waiting to meet the candidates elected by the majority of s as this speeds up the process by automatically and immediately counting votes.

Among the advantages of using electronic voting we have the following presented in Figure 3.

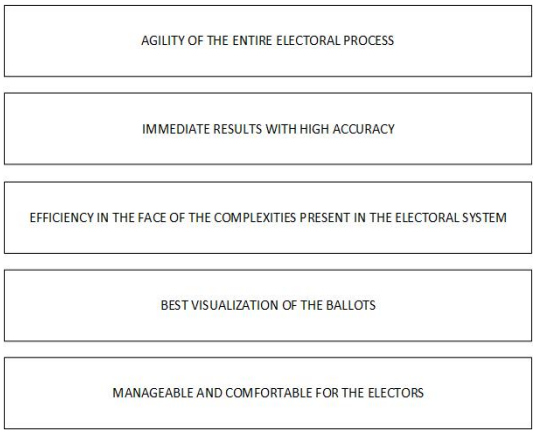


Figure 3. Advantages of using Electronic Vote

III) Electronic Vote Controversy

Despite the many benefits of the electronic voting system, doubts remain around it, in some parts of the world.

Many of these doubts arise due to the lack of knowledge that the company still has in relation to the functions of the electronic voting system, its use, access, among others. Another doubt that society maintains is above all what is related to the security of the system, that is, whether the system effectively keeps the vote, maintaining its integrity, showing no kind of relationship between the vote and its voters.

A. Possibility of System Infringement

Electronic voting systems have a number of measures to regulate or reduce vulnerabilities in the use of these software. Author Lagunez notes that political activity on the Internet is not just marketing, nor social mobilizations; governments have been violated over the Internet by hacker groups called cyber-activists, ideological groups operating over the network. the political activity of these groups goes beyond symbolic protest, they are able to infiltrate the official pages of governments and political, religious and even criminal groups. The space of political struggle becomes more significant because it is shown that the state does not have complete control of a space that can be used by many people; although it is important to remember that it is not yet a totally global phenomenon [11].

Therefore, this modality cannot be considered very reliable either. On the other hand, even if it does not hurt to consider some points that in some way or another can help them not happen so often and do not cause any alterations.

The author Castillejos Lopez indicates that addressing the issue of security in digital environments invites us to reflect on the benefits of using the internet in 21st century society. [12]. The author Bolañoz Burgos points out that while Ecuador defines computer law as the set of legal systems established in order to regulate the processing of information [13].

These vulnerabilities can occur at any time, in an unforeseen way in such a way that the information that is stored can be manipulated and in turn disclosed which could have various negative consequences, in addition to alter the results of these elections.

B. Actors Who Can Break the System

These vulnerabilities are commonly presented by third parties, either with some malice for committing fraud or other evil, nor is it ruled out the fact that this can happen due to ignorance of it and therefore has not been Intentional. The author García, says that security in computer networks will always be a relevant and of interest topic at the computer level [14]. So also the author Castillo points out that precisely because of the different vulnerabilities that exist is that not all countries implement those electronic voting systems for fear of their alteration or fraud that may exist, even because of this uncertainty if your data and your choice will be secure and alter your integrity [15]. This weakness is latent and can affect the security, integrity, availability of sensitive information these systems handle.

In this way, it will be taken into account that these failures do not always arise specifically by electronic voting as such, but that there may be multiple causes such as in the architecture of the system in such a way that it may cause some alteration of the votes made.

Author Avila mentions the campaign season providing voters with a complex information environment in their search for information to guide their vote. These vulnerabilities can become threats to the good performance of electronic voting systems, as environmental and natural threats in addition to the human threats already discussed above. These threats can affect the electoral process and alter the results. Many of these vulnerabilities and threats are uncontrollable by humans as they can happen

spontaneously, however, certain situations can be prevented. The author Lavin José points out that, once these tools and technology have been examined, it can be said that there is a good electronic basis for establishing a Model of Deliberative Democracy since we not only have tools that would allow the vote, but also also the debate, reflection and weighting of opinions [17].

C. Suggestions for Decreasing System Infringement

It is vitally important that society's opinion on the implementation of new voting arrangements is reported and considered as all citizens must feel confident in the electoral process so that this does not then influence the decision of their vote. Keeping voters informed will reduce vulnerabilities to the system, as everyone will cooperate with the care of their votes and follow protocols in every part of the electoral process.

So citizen participation must be a constant reality, in this way it can be said that it has been worked democratically, then each voter will be able to have the pleasure of accessing their right to vote according to the candidate who meets their expectations, because is a manager of the electoral achievements of your city or country.

Thus, the author Marquez argues that the electoral behaviour of the citizen is based on the calculation of the expected usefulness made by the voter, taking into account the proposals presented at the particular juncture, weighing the costs and benefits of each [18].

IV) Information Systems Security

A. Cryptography

One of the recommendations in the application of electronic voting is the use of Public Key Cryptography Encryption, as expressed by author Marin Bermudez, In the context of telematicvoting systems, public key cryptography is often used to encrypt the vote, so that only the holder of the private key can access and account for it:

- In the previous phase of preparing the vote, the key pair is generated; public and private.
- The public key is distributed to all voters to encrypt their vote with it.
- The private key is kept by the board or electoral committee. Generally this key is chopped and distributed among the members, so that a single member cannot access to decipher the votes.

Once voting is complete, members of the electoral board or committee enter the private key so that they can decipher the votes and recount. This ensures access to voting data [19]. Once voting is complete, members of the electoral board or committee enter the private key so that they can decipher the votes and recount. This ensures access to voting data [19].

A. Event Log

Public institutions are also concerned with the need to introduce other security measures into electoral processes, as author Vera Victor says that there are companies that do not care about implementing computer security measures or if they do they only consider externalities and do not take into account the risks that may arise within them [20].

The author Morales points out that in electronic voting systems it is extremely important to have a record of events, in order to have the evidence of actions taken especially for cases where manipulation is suspected. in a voting session, for example, you can generate the registration of voter authentication, the event of the selection of candidates, the closure of the session, etc. [21].

B. Preventive Security

Any media or information system must have security requirements and actions to be taken in the presence of a vulnerability or threat. To this end, the following aspects are proposed:

- Prevention of security attacks, which refer to the actions that will be carried out if there is any violation of the information.
- Security mechanisms, which refer to the actions that will be carried out in order to identify, prevent and block any attempt to attack stored information.
- Security services, these refer to the service provided to the person when storing their data and other information.

C. Voting Subject Check

Several electoral contests have shown in the list of voters to citizens who cannot exercise their right to vote, because they have died or there have been candidates in two ballot boxes, which can affect the results, therefore, a secure system is advisable, which it would involve the on-site verification of biometric data, for example, the finding of fingerprints by scanning them to confront those registered at the time of entry to the register, or the iris record of the subject who shows up to vote. This voter check would lessen attacks by hackers who want to use these non-existent voters and turn them into real votes. The author Del Blanco points out that the growing proliferation of attacks makes proper system maintenance crucial. On the other hand, no one can predict long-term computational capabilities [22].

D. Voting Subject Check

For the voting system to achieve its objectives, honest participation of political parties and government is needed, as democracy must be embodied in concrete actions that candidates must respect and maintain. This is the opinion of author Lopez that, in this way, political efficiency would be a neutral element that would call for a vote, but which is earlier or higher than mere voter participation, which is precisely what interests us [23]. To maintain a secure voting system, it is necessary to incorporate certain requirements into work, according to the author Valencia indicates that democracy in the 21st century cannot be limited to the right to vote and be elected. The complexity of our societies and the existence of a plurality of legitimate interests make it necessary to expand the spaces of participation and control in the various entities of the State [24]. On the other hand, the legal framework will allow to sustain every decision that is made around the elections within the country, as mentioned by the author Montalván Calderón, which it is also important to establish certain

guidelines that help to meet legal requirements that have electronic security, these encompass various procedures, transparency, verification and reliability [25].

3. Results

Several results could be achieved in this study:

- Security Analysis in electoral systems in Latin America, according to Hosp and Vora's theorem, is vulnerable and requires greater security measures.
- Suggestions to follow in electoral security systems, with the respective applicability in electronic voting.

In addition, for greater security in the process, a probabilistic model of passwords is chosen: A probabilistic model of passwords is determined by any function P, defined in the space of possible passwords (S) in the range [0.1], which assigns a P(s) to each password in such a way that:

- P(s) ≥ 0, for all password $S\ e^{S}$
- Σ

These models are a critical tool for investigating password security. the definition and interpretation of P(s), depends on the model. The existence of alphanumeric password databases available on the Internet has allowed experimental capture of its characteristics, which are used for the definition of P(s), Markov strings and context-free grammars have been the two models most used to quantify by (P) the probability that password s will be selected, by the user, in the registration phase. These P(8) values are the basis for dictionary attacks and some password security assessment metrics [26].

So also, when we mention the Encryption System, we have El Gamal. Originally ElGamal has the multiplicative homomorphic property, but for the vote count a slight variant is used that is additive homomorphic and that is the one that allows the sum of encrypted votes:

Be a G-switched group of order |G| = q. The public key is (G, q, g, h), being "g" a generator of G, $h = g^{x}$,. "x" it's the private key.

Sea in the encryption of a vote v

$E(v) = (\alpha, ß) = (g^{r}, p^{v}, h^{r})$

Being "r" a random value $r \in \{0,1, \ldots, q - 1\}$ and "p" another G generator independent of g.

For two votes v_1 and v_2 encrypted such as:

$E(v_1) = (\alpha_1, ß_1) = (g^{r1}, p^{v1}, h^{r1})$

$E(v_2) = (\alpha_2, ß_2) = (g^{r2}, p^{v2}, h^{r2})$

The additive homomorphic property is:

$E(v_1) \cdot E(v_2) \cdot (\alpha_1, ß_1) \cdot (\alpha_2, ß_2) = (g^{r1}, p^{v1}, h^{r1}) \cdot (g^{r2}, p^{v2}, h^{r2})$

$= (g^{r1+r2}, p^{v1+v2}, h^{r1+r2}) = E(v_1 + v_2)$

Among the advantages of VER systems that use EEH are: when operating on encrypted votes you do not need an anonymous channel (unlike the blind signature scheme), the counting process is very efficient since you don't have to decrypt the votes one by one before counting and you don't have to wait for the polls to start the polls to begin counting [27].

A. *Security Analysis in electoral systems in Latin America, according to Hosp and Vora's theorem, is vulnerable and requires greater security measures.*

According to Hosp and Vora's theorem, there is no electoral system with perfect integrity, verifiability and privacy, as malicious attempts to modify the results will always be found. But you can count on important suggestions such as individual reinsurance, which does not allow to reveal the identity of the voter, as well as transparency, that allows access to the result obtained, the appropriate use of technological equipment that does not store voter registration and also restrict access to the different equipment used.

B. *There are different ways of violating the system, internally and externally, for which they need to be identified and followed by suggestions from the team responsible for electoral protection. Suggestions to follow in electoral security systems, with the respective applicability in electronic voting.*

Several Latin American countries have implemented security measures and considered the security measures that are applied in the votes of other countries in the region, while the will to implement electronic voting is being considered, including Argentina, Chile, Peru, Uruguay and Costa Rica. In the region only Brazil and Venezuela show the use of electronic voting, but that, as in any electoral system, there are reports of the results obtained. To pre-caution voting in a country, using new technologies, they should consider one of the three options being used: automated reading of paper voting, autonomous electronic voting and networked electronic voting.

Hence to take relevant security measures, such as preventive security that takes care of the whole process, cryptography, with access keys, the log of events that allows to audit each step, before, during and at the end of the respective vote, the vote and all this in an environment that has been cared for under specific law articulations, in which citizens have participated, as well as political parties.

Hence, it is suggested to work on whatever is necessary, as figure 4 points out, which motivates intervention in all the spaces relevant to the electoral contest.

4. Discussion

Implementing a security model is an important requirement for the use of an electronic voting system, as has been emphasized above, since any electronic system in itself does not provide 100 percent security and because every process is finds susceptible to vulnerabilities and threats that can be raised at any time.

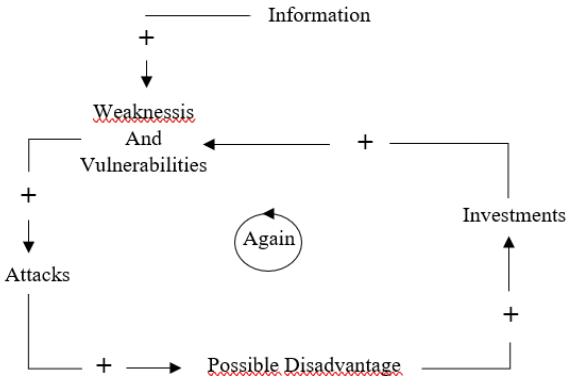


Figure 4: Process to care for a vulnerable system

Any attack on the electoral system will make an impact, but it is possible to reduce it to maximum expression, so you have to be prepared to attack them as soon as they are identified. It is therefore imperative to train staff so that they can follow technical suggestions as needed. In addition, to complement this good management and development of the electronic voting system, it is important to consider that society needs to know the system to be used, its benefits and use. This should be affordable to everyone and must ensure the trust of each of them.

5. Future Work and Conclusions

A. Future Works

- Electronic voting systems with their respective security, for voters who cannot mobilize to their polling station.

- Design of protocols and processes with latest technological tools, for public institutions that control voting in the country.

B. Conclusions

The growth of technology and its easy access has allowed it to be used in different areas and politics is no exception, which is why different modalities of electronic voting have been implemented so that they can reduce some failures that are maintained in the traditional voting system. Some Latin American countries have already implemented and others want to implement several of these models, but in turn it is important to consider a security model that ensures that reliability and integrity of the secret election vote, bearing in mind that it is not equally guarantees its total security, but it does a way to reduce and prevent certain failures.

Latin America suffers from corruption problems and popular voting is always threatened by subjects who wish to change the results, so that the more security options exist during the protocol, the more optimal results can be obtained.

Therefore, it is important to work with security measures that take care of the decision of each voter, such as preventive security, cryptography, auditing, together with the articulated law that are stipulated in the country.

Acknowledgements

The authors thank the Salesian Polytechnic University of Ecuador, the research group of the Guayaquil Headquarters "Information Technology, Security and Information for a Globalized World" (CSITGW) created in accordance with resolution 142-06-2017-07-19 and the Secretariat of Higher Education Science, Technology and Innovation (Senescyt).

References

[1] Diaz-Ricardo, Yanet; Perez-Del Cerro, Yunetsi; Proenza-Pupo, Dayami. System for the Management of Computer Security Information at the University of Medical Sciences of Holguin. Holguin Sciences, 2014, vol. 20, no. 2, p. 1-14.

[2] Mube, Mirna, Rivas, Lizbeth. Current status of computer security incident response teams. RISTI-Iberian Journal of Informacao Systems and Technologies. 2016, No. SPE3, p. 1-15.

[3] Curious, Walter H.; Espinoza-Portilla, Elizabeth. Conceptual framework for strengthening health information systems in Peru. Peruvian Journal of Experimental Medicine and Public Health, 2015, vol. 32, p. 335-342.

[4] Jerez, Ariel; Maceiras, Sergio D'antonio; Maestu, Enrique. Public spheres, political crises and the internet; Podemos's electoral emergence. History, Science, Saúde-Manguinhos, 2015, vol. p. 1573-1596.

[5] Rovelli, Laura-Ines. Recent expansion of prioritization policy in scientific research of public universities in Argentina. Ibero-American Journal of Higher Education, 2017, vol. 8, no. 22, p. 103-121.

[6] Ottoboni, Kelly; Stark, Philip. Election INtegrity and Electronic Voting machines in 2018 georgia, USA, E-Vote-ID 2019 Proceedings, 2019.

[7] Mellinghoff, Rudoff. The German electoral system: overview and new trends. Journal of Electoral Law, 2014, no. 17. p. 1.

[8] Schimidt, Jeff, Alfaro, Jaime Gutierrez. Towards the development of a prototype electronic voting system for Costa Rica. Technology in Progress, 2016, vol. 29, No. 3, p. 146-158.

[9] Sandoval Ballesteros, Netzai. jurisprudential admission of internet voting for foreign residents of Mexico City. Mexican Bulletin of Comparative Law, 2015, vol. 48. 142, p. 275-312.

[10] Abdala, Mary Bethlehem; BENESCH, Pedro A. Antenicci. Evaluation of the effects of the Single Electronic Ballot: experimental evidence of the elections in Chaco 20156. SAAP Magazine. Publication of Political Science of the Society. Argentina political analysis, 2016. vol. 10, No. 2, p. 339-354.

[11] Lagunez Lopez, Oscar Nicasio. for effective democracy: X-ray of a stagnant political system, 1977-2012. Region and Society, 2014, vol. 26, No 59, 9,293-301.

[12] Castillejos Lopez, Berenice, Carlos Arturo Torres Gastelú, and Agustín Lagunes Domínguez. "Security in the digital skills of millennials." Opening (Guadalajara, Jal) 8.2 (2016): 54-69.

[13] Bolaños-Burgos, Francisco; Gomez-Giacoman, Christopher. Qualitative study of the relationship of laws and computer expertise in Ecuador. Receives. Electronic Journal of Computing, Computing, Biometrics and Electronics, 2015, No. 3.

[14] Garcia, Paulo Gaona; MARON, Carlos Montenegro; VELANDIA, Julio Baron. Ontological model for predicting computer attacks from virtualized Honeynets. Logos, Science & Technology Magazine, 2016, vol. 8, No. 1, p. 101-114.

[15] Castillo, Jessica Nataly, et al. Model for reducing it security risks in web services. Summits, 2018, vol. 4, No. 2, p. 19-30.

[16] Avila, Caroline; Cabrera, Gabriela. the effect of the rumor on the change in voting: the anger, fear and uncertainty generated by the rumor in electoral processes and its contributions to voter decisions. Sign and Thought, 2016, vol. 35, no. 69, p. 100-116.

[17] Lacon, José M.; Alvarez, Edison; MAJOR, Franklin. Deliberation and participation: electronic trails. meth. social sciences magazine. 2014, vol. 2 No. 2.

[18] Marquez, Stefany Arteaga. How does the political brain work? guide to political communication to understand voters and public opinion. Mario Alario D'Filippo Magazine. 2018, vol. 10, no. 10, p. 187-212.

[19] Maron Bermudez, Antonio. Study of the use of blockchain protocols in electronic voting systems. 2016.

[20] Vera, Victor Daniel Gil; Vera, Juan Carlos Gil, Organizational Computer Security: a simulation model based on system dynamics. Scientia et technica, 2017, vol. 22 No 2, p. 193-197.

[21] Rocha Morales, Victor; Ruiz Hernandez, Oscar; Fernandez Martinez, Luis Felipe. Audit mechanism for the detection of vote tampering in electronic voting systems. PAAKAT: Technology and Society Magazine, 2018, vol. 8, No. 14.

[22] Of White, David Y. Marcos; ALONSO, Luis Panizo; ALONSO, JA Hermida. Development of an advanced methodology for evaluating Remote Electronic Voting Systems. 2016.
[23] LOPEZ, Magdalena. "Democracy in Paraguay: "The interruption of the process of change" with the dismissal of Fernando Lugo Méndez (2012). CENDES Notebooks, 2014, vol. 31, no. 85, p. 95-119.
[24] Valencia Tello, Diana Carolina; Chueiri Karam, Vera. Accountability, Reporting, And Management Control. How Does This Work in Argentina According to the legal System in Force? Legal Opinion Magazine, 2016, vol. 15, no. 29, p. 165-185.
[25] Montalvon Calderon, Aricka, et al. E-vote in the field of local administrations 2015.
[26] Legon, Carlos Miguel, et al. New probabilistic model in graphic authentication. ISSN Electronic, Automation and Communications Engineering Magazine: 1815-5928, 2019, vol. 40 no 3, p. 92-104.
[27] Of White, David Y. Marcos; Alonso, Luis Panizo; Alonso, JA Hermida. Development of an advanced methodology for evaluating Remote Electronic Voting Systems. 2016.

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 61-68 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Mechanical Testing Methods for Body-Powered Upper-Limb Prostheses: A Case Study

Renato Mio[1,*], Midori Sanchez[2], Quino Valverde[3], José Lara[4], Francisco Rumiche[4]

[1] *Laboratory of Biomechanics and Applied Robotics, Pontifical Catholic University of Peru, 15088, Peru*

[2] *Digital Manufacturing Laboratory - VEO 3D, Pontifical Catholic University of Peru, 15088, Peru*

[3] *INACOM Group / Aula PUCP-CIMNE, Pontifical Catholic University of Peru, 15088, Peru*

[4] *Mechanical Engineering Section, Engineering Department, Pontifical Catholic University of Peru, 15088, Peru*

ARTICLE INFO

Article history:
Received: 31 May, 2019
Accepted: 18 August, 2019
Online: 04 September, 2019

Keywords:
Upper-limb prosthetics
Material testing
3d-printing
Benchmark testing
Body-powered prostheses

ABSTRACT

3D-printing technologies have greatly influenced the field of fabrication of medical devices. In particular, Fused Deposition Modeling 3D printing has emerged as one the most popular and most promising technologies for fabricating upper-limb prostheses. Over the last years, a variety of types and designs of 3D-printed hand prostheses have been created and are commercially available. However, there are no standards or established procedures for testing these devices. Available information regarding their long-term performance and functionality is very limited. This paper presents a case study of mechanical testing methods applied to a specific design of an upper-limb prosthesis. The device and its subassemblies were subjected to flexion test in hyperextension and abduction conditions, fatigue/wear test, and tensile test. The experimental results are presented and examined. Testing procedures, adaptations and recommendations are described and discussed to demonstrate ways of generating reliable data that serve for comparison among different hand prostheses designs.

1. Introduction

The fabrication of medical devices is one of the most influenced fields by 3D printing technologies advances. Particularly, in the last years, 3D printed hand prostheses have emerged as a promising low-cost alternative, compared to conventional hand prostheses, to fulfill the need for these devices in developing countries [1]. Currently, Fused Deposition Modeling (FDM) is one of the most common and inexpensive technologies for fabricating functional end-use parts by 3D printing, including hand prostheses. However, FDM has a disadvantage related to the anisotropy generated by the process and its corresponding effect on mechanical properties. A material fabricated by 3D printing shows a ductile behavior in the orientation of the bead deposition, but a brittle one in the perpendicular orientation. Therefore, FDM printed components are susceptible to fracture in the building orientation; a failure mechanism called delamination. While the effect on mechanical properties of many fabrication parameters such as layer height, infill pattern, density, nozzle diameter, build plate temperature, printing orientation and velocity [2, 3, 4, 5, 6, 7] has been evaluated using 3D-printed test specimens, little advances have been made in testing and characterizing entire components fabricated using FDM 3D-printing.

Although the popularity and the use of 3D printed hands keep increasing, there is a lack of standards for fabrication and testing of these devices and very limited information available in terms of their long-term performance and functionality. Existing hand prostheses standards and testing procedures are not directly applicable, even for currently commercial multi-fingered and dexterous hand prostheses, because they were conceived mainly for conventional hook prostheses. Moreover, hand prostheses manufacturers use their design-specific tests to provide full specifications of their commercial products.

In this context, different researchers have explored the possibility of adapting existing methods and implementing variations, according to their investigation's objectives and available resources, to evaluate the performance of open-source and specific models of 3D printed upper-limb prostheses. Tests have been conducted to evaluate mechanical properties such as tensile strength, impact resistance, flexural strength, and fatigue

* Renato Mio & Email: r.mioz@pucp.edu.pe

resistance, both in the whole prostheses and sub-assemblies, with parameters (test velocity, applied load, number of cycles, etc.) that resemble conditions of activities of daily life (ADL). A review of these tests has been presented by the authors in the conference BIBE 2018 [8] and is summarized as follows.

Tenim [9] and Koprnicky [10] analyzed the mechanical behavior of phalanges of prosthetic hands employing hyperextension (see Figure 1a) and lateral bending test and cyclic flexion-extension test (see Figure 1b). The bending tests were intended to evaluate the maximum load that a phalange can withstand and extrapolate the results to compute the maximum load that can be carried by all of the fingers of a prosthetic hand. These bending tests have been also applied to individual fingers from robotic grippers [11]. Furthermore, the purpose of the cyclic flexion-extension test was to measure the endurance of a finger under the application of several fatigue cycles and to determine the mode of failure under this condition (e.g. damage of one of the finger components or wear of the flexion cables channels). In addition, Crisco and Wolfe [12] analyzed the finger biomechanics and proposed testing procedures to mimic the conditions to which a prosthesis is subjected to when conducting activities of daily living. While the ASTM F1781 standard [13] provides some guidance for finger fatigue test, Joyce and Unsworth [14] have proposed a set of parameters to be applied, such as the flexion-extension load magnitude: 10-15 N, pinch load: 100 N, the angular amplitude of the motion: 0-90 degrees, and the frequency: 1.5 Hz.

Among the tests to which prosthetic hand subassemblies could be subjected to, the ISO 22523 standard [15], annex A, proposes the distal tensile test (see Figure 1c). In this test, a hand subassembly is induced to perform a power grip on a 19 mm diameter cylindrical object. As schematically observed in Figure 1c, the object is pulled away, forcing the fingers to extend or to break. In this way, it is possible to establish the maximum load that the prosthesis can withstand without failure. Additional tests involving grasping movements with a cylindrical object have been proposed in [16, 17, 18]. Cylindrical grip test is of key importance because it mimics the movement conditions of most of the body-powered hand prostheses, i.e. opening and closing of all of the fingers at the same time. Furthermore, the ISO 22523 standard describes the procedures to conduct hand or wrist flexion tests in the case of transradial prostheses.

This paper presents a case study of mechanical testing methods applied to an upper-limb prosthesis. The study is focused on the 3D-printed D1M Hand (acronym in English for Dando Una Mano, *Giving a Hand*) [19] (see Figure 2). The D1M hand was designed for transmetacarpal or partial hand amputations and was fabricated by the FDM 3D printing process. Due to its design and functionality the D1M hand was subjected to finger flexion test (in the hyperextension and abduction/adduction directions) to measure individual finger maximum load, finger cyclic flexion-extension to measure individual finger durability, and hand tensile test to measure the maximum possible load that the hand can withstand while performing a cylindrical grasp. Schematic drawings and a description (references, conditions, outcomes, and objectives) of these tests are shown in Figure 1 and Table 1 respectively.

This work complements the one presented by the authors in BIBE 2018 [8]. This manuscript includes a summary of the design and fabrication process of the D1M hand, the experimental details and testing procedures description; with particular emphasis placed on the machine adaptations for positioning and supporting the tested hand and finger samples, and the discussion of obtained results.

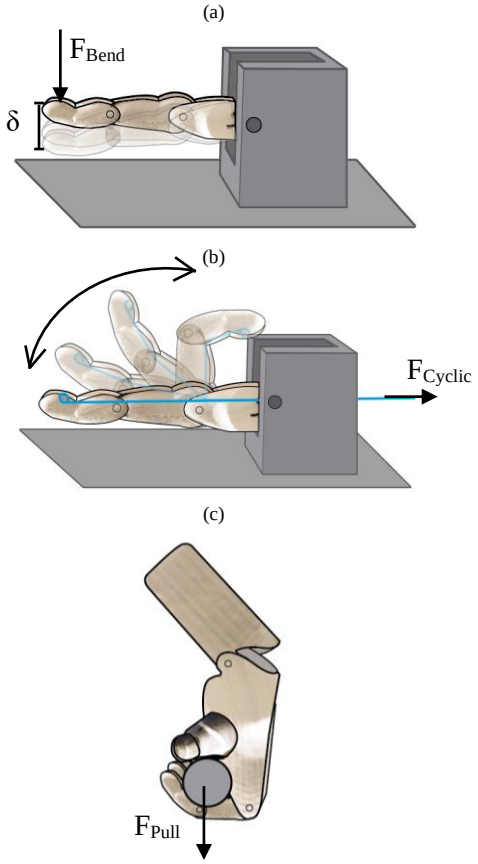


Figure 1. Schematic drawings of the different mechanical tests found in the literature. (a) Finger flexion test, (b) endurance/fatigue test and (c) hand tensile test. Adapted from [8].

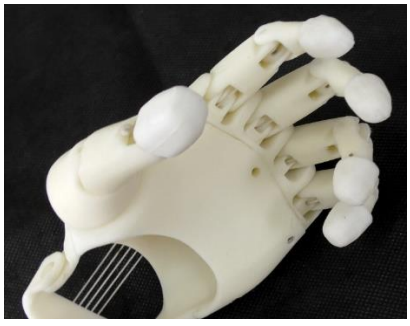


Figure 2: D1M hand prosthesis prototype. Image from [19].

2. Experimental Procedure

The mechanical tests were performed using universal testing machines. The D1M hand was fabricated using a commercially available desktop 3D printer. The use of commercially accessible equipment will allow the experimental procedures to be replicated and the results to be compared with those obtained by other researchers. The assemblies and accessories presented in this work can be modeled in CAD software and then minor modifications can be made to accommodate different designs. Validation of the printing parameters and methodology was performed with detailed testing applied to the D1M hand.

Table 1: Description of the mechanical resistance tests applied to the D1M Hand (adapted from [8]).

Type of test	*Test description*			
	Based on (guide or standard)	***Test conditions***	***Recorded outcomes***	***Objective***
Finger flexion	ISO 178, ASTM 790 and the works of Tenim [9] and Koprnicky et al. [10]	Velocity of the test: 1% the gage length per minute. Finalize when: Maximum deflection is 5% or failure. Different tests: Hyper-extensive and lateral loading.	Displacement and force against time.	Estimate the mechanical strength of the fingers to withstand forces that over-extend (hyper-extensive loading) or bend the fingers laterally (lateral loading). The target load value is 30 N according to [20]
Finger cyclic flexion-extension	ISO 14243-1 and the works of Tenim [9] and Joyce et al. [14].	Cyclic loading frequency: 1.5-2 Hz. Finalize when: achieve 1,200,000 cycles or failure of whichever component. Load at the MCP joint: 10 - 15 N. Pinch force: 100 N. Finger base joint range of motion: 0 - 90 degrees.	Cycles can achieve the prosthesis until the failure. Wear of the components measured by weight difference before and after the test.	Calculate the number of cycles the prosthetic finger can withstand. The target value is 1,200,000 cycles (equivalent to 4 years of use) according to [21]
Hand tensile test	ISO 22523-2006, annex A.8.2 [15]	The prosthesis grasps a 100mm x diameter 19 mm cylinder. An external force pulls the cylinder out of the hand. Preload: 10 N Load: increases at a rate of 1 – 10 N/s until failure.	Displacement and force against time.	Estimate the maximum payload of the whole prosthesis when performing a cylindrical grasp.

2.1. Fabrication of the D1M hand prosthesis

The D1M hand was fabricated by FDM 3D printing with an Ultimaker 3+ Extended machine; the material employed was acrylonitrile butadiene styrene (ABS). Table 2 shows the general 3D printing process fabrication parameters that were applied to all of the main components (fingers, palm and forearm).

The prosthesis was designed based on the anthropomorphically structure of the user and the functionality of the 4 linkages mechanism [19]. ABS is an optimal polymer for this use because of its strength and low density. As mentioned previously, the mechanical resistance of the 3D printed hand depends on the manufacturing orientation, caused by the intrinsic anisotropy of the manufacturing and its deposition of consecutive layers. For this reason, optimal printing orientation for each component was chosen based on factors such as maximizing the area of the critical cross-section under stress (to prevent delamination) and attaining the best surface quality with no support material for important geometrical features (like shaft supports). Table 3 and Figure 3 shows the orientation of each component of the hand and its printing parameters. In addition, it was observed that for the palm less infill density is placed than in the fingers, because in the continuous use of the prosthesis it was identified that the fingers were critical pieces, due to the small thickness in the lace. Many other existing hand prostheses designs also come with suggested infill percentages, which adds another variable influencing the resulting mechanical properties of the resulting device. With the purpose of finding the exact technical specifications of the prosthesis, mechanical tests were performed based on the experimental parameters found in the literature.

2.2. Experimental Procedure on D1M hand

Three different tests were conducted in the D1M hand: first the hand tensile test, chosen because the main actions performed by upper limb prostheses require a cylindrical grip; then, finger flexion tests were performed in order to know the maximum weight that an individual finger can resist; and finally, the finger fatigue / wear test to estimate the prosthesis' durability and the wear due to its use. These tests required the design of custom fixation components to be placed between the hand components and the testing machine.

Table 2: 3D Printing parameters for the components of the D1M Hand.

Parameter	**Value**
Nozzle Temperature	230 °C
Build plate Temperature	100 °C
Printing speed	65 mm/s
Layer height (Z-axis resolution)	0.2 mm
Wall thickness	1.2 mm
Infill pattern	Linear

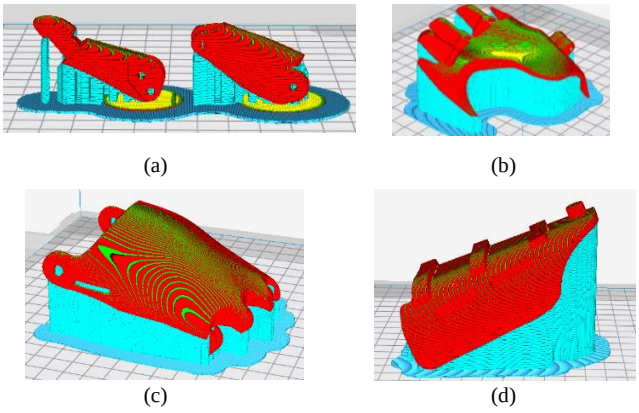
Table 3: 3D Printing Parameters specific to each component.

Piece	**Layer thickness [mm]**	**Infill [%]**	**Angle orientation**	**Figure**
Finger	0.1	50	17° and 25°	Figure 3a
Palm (front side)	0.2	30	28°	Figure 3b
Palm (backside)	0.2	30	15°	Figure 3c
Forearm	0.2	50	32°	Figure 3d

2.2.1. Hand tensile test

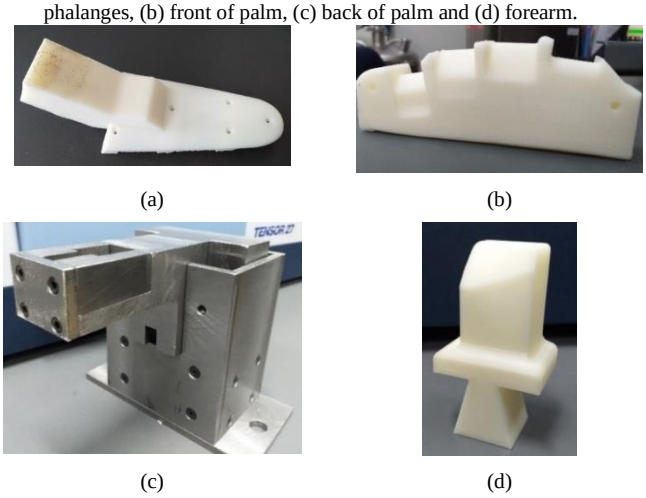
The tensile test was performed in the whole hand assembly to evaluate the maximum load that the prosthetic hand can withstand when it is performing a cylindrical grasp movement. The test was conducted using a Zwick Roell Z050 tensile test machine with the test parameters defined by ISO 22523:2006 [15]. Three whole hand assemblies were tested. As shown in Figure 4a, a flat area accessory to connect the machine jaws with the hand was fabricated, it also served to keep the assembly aligned with the axis on which the machine exerts force. This accessory was printed in ABS and 100% filling density was used. Additionally, a grasping device (see Figure 4b) was fabricated to allow the prosthesis to make a cylindrical grip. On the other hand, a support device was added to maintain the wrist flexion needed to close the fingers around the grasping device. To ensure that the test can reproduce a real-use condition, the prosthesis must be tested while

activated with its fingers closed. For this reason, a pair of accessories were designed to keep the prosthesis locked as if it were activated by a user exerting force with the stump. A steel accessory (see Figure 4c) was manufactured to support the hand and withstand relatively high loads without deforming. In addition, an accessory acting as a support point of the inner part of the palm was printed (see Figure 4d). Figure 5 shows the complete assembly to perform the testing.



(a) (b) (c) (d)

Figure 3. Orientation of the hand prosthesis' main components: (a) finger phalanges, (b) front of palm, (c) back of palm and (d) forearm.



(a) (b) (c) (d)

Figure 4. Accessories for the optimal fix between the prosthesis and the testing machine.

The test is performed until the failure of any component (or until the prosthesis stops holding the grip due to the deformations that occur in the threads that cross the phalanges).

2.2.2. *Finger Flexion test*

Flexion tests were conducted in the fingers to evaluate their maximum load-carrying capacity. The tests were carried out simulating two conditions: a finger extended in a horizontal plane (hyperextension test), and a finger extended in a vertical plane (abduction test). The tests were performed using a Zwick Roell Z0.5 multi-test machine. Three samples of middle and little fingers (longest and shortest finger of a hand, respectively) were subjected to each condition of the test. The test speed was proportional to the length of the finger, which in turn depended on the test condition. For the hyperextension condition, the test speed was 6.6 mm/min and 4.6 mm/min for the middle and little finger, correspondingly. For the abduction condition, the speed was 7.3 mm/min and 5 mm/min for the middle and little finger respectively. Three accessories were needed for this test: the general coupling, the secondary coupling and the bending punch. Figure 8a shows the assembly designed in the CAD software. The general coupling (see Figure 6) was fixed on the machine table and housed the secondary coupling and the finger. The secondary coupling (see Figure 7) varied depending on the size of the finger and the type of test (hyperflexion or abduction). The bending punch was fabricated from a sheet of steel and had a U shape in order to avoid the slip with the 3D printed finger, as shown in Figure 8b.

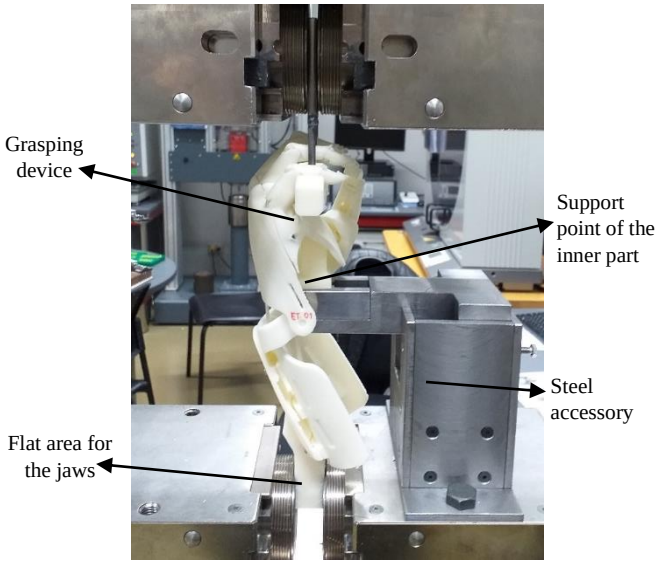


Figure 5. The complete assembly of the hand tensile test. Arrow indications correspond to the accessories shown in Figure 4.

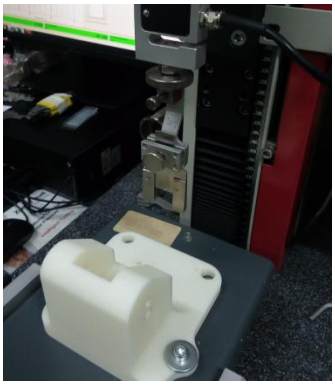


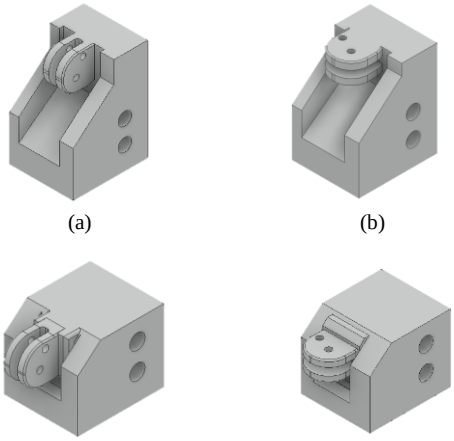
Figure 6. 3D printed coupling made of white ABS and punch that exerts the bending force.

2.2.3. *Finger fatigue/wear test*

The fatigue/wear test was performed to evaluate the capacity of the index finger assembly (3D printed phalanx, wire and pins) to perform a cyclic pincer grasp movement and measure the possible wear between the threads and the internal channels of the prosthesis, and the metal pins with the phalange holes. A Zwick Roell Z0.5 machine was used for the test, subjecting samples of the index finger to 24000 cycles in flexion mode. The test load variation was from 2 N to 80 N to ensure the force in the fingertip of 15N (as suggested in [21]), with a displacement speed of 1000 mm/min. The wire that pulls the mechanism to operate the finger was placed in the machine's jaw (see Figure 9) and was connected with the finger through a bronze pulley (see Figure 10b) that also that also serves as an alignment tool. Then the finger was set up in

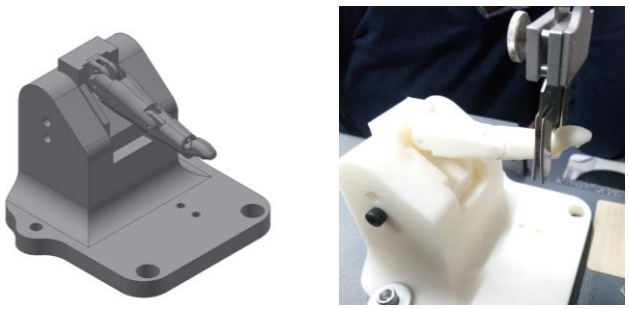
a 3D printed holder to facilitate the test and a sensor (a Force Sensitive Resistance sensor with a magnitude range: 100 g – 10 kg) was positioned in the area of contact with the fingertip in order to evaluate the pincer grasp load (see Figure 10a).

Three samples were tested and, besides analyzing their capacity to withstand cyclic loads, the wear resistance of the assembly was evaluated by weighing the samples before and after the test.



(a) (b) (c) (d)

Figure 7. Accessories designed for the (a) middle finger hyperextension, (b) middle finger abduction, (c) little finger hyperextension and (d) little finger abduction tests.



(a) (b)

Figure 8. (a) CAD design and (b) fabricated components of the finger flexion test assembly.

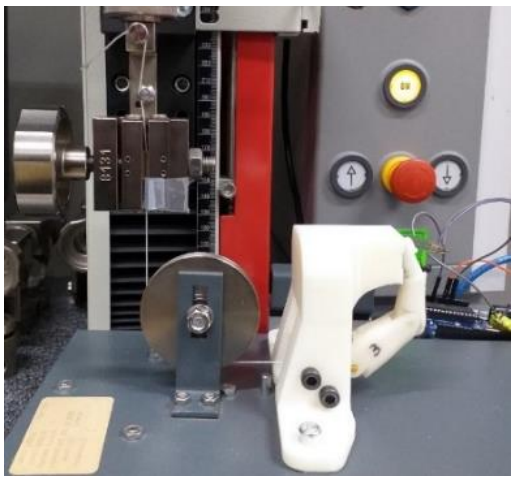
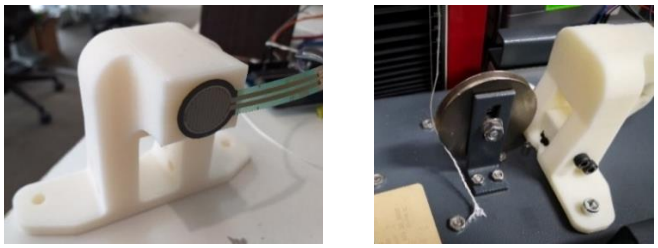


Figure 9. The complete assembly for the fatigue/wear test.



(a) (b)

Figure 10. Accessories for the finger fatigue/wear test: (a) sensor on its 3D printed holder and (b) pulley.

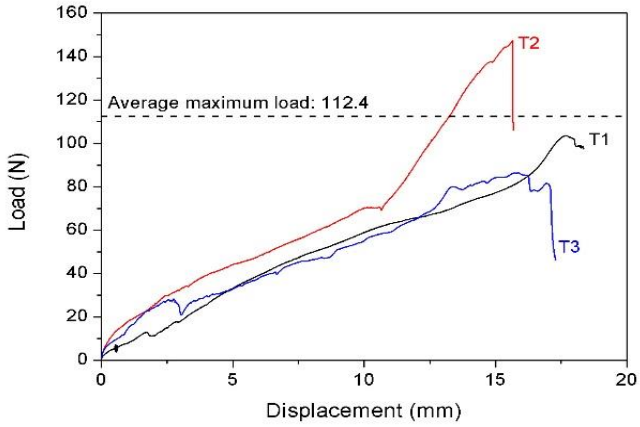


Figure 11: Tensile test curves for the D1M prosthetic hand samples

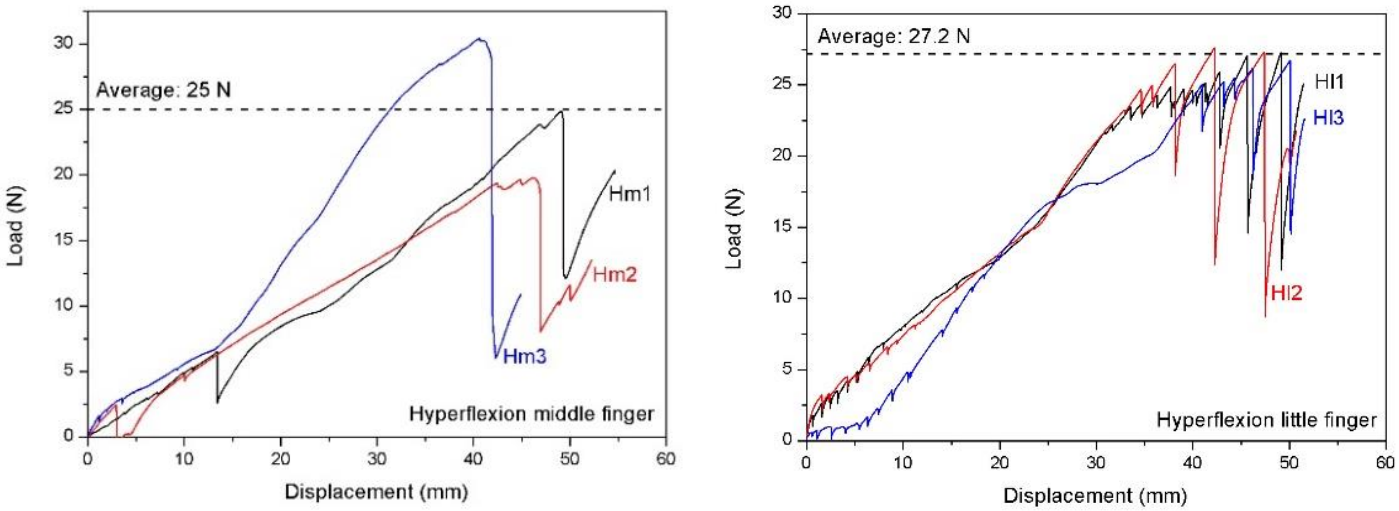
3. Results and Discussion

3.1.1. Hand tensile test

Figure 11 shows the load-displacement curves for the tensile test in the three assembled prosthetic hands (samples T1, T2 and T3). The average maximum tensile load that the hand can withstand before failure is 112.4 N. Discontinuities in the curves and differences in behavior for each sample could be attributed to factors inherent to the 3D printing process, and to re-accommodation events of the hand assembly through the test due to the elastic and plastic deformations experienced by the wire.

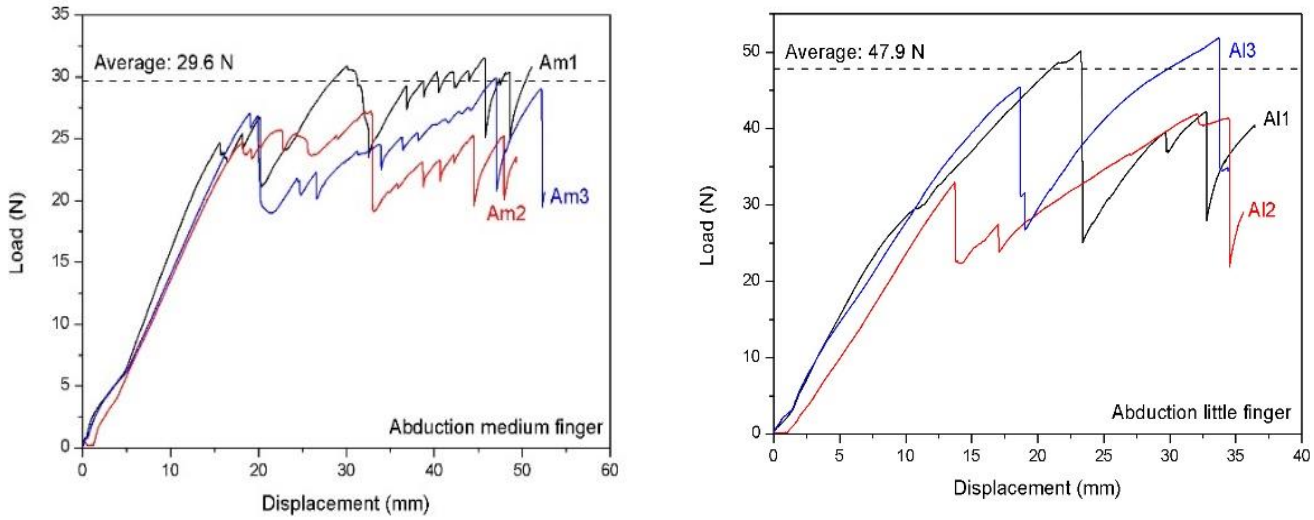
3.1.1. Finger Flexion test

Figure 12a and Figure 12b show the load-displacement curves for the middle and little finger samples obtained from the flexion tests in hyperextension condition (hyperflexion). The average maximum load for the middle and little finger in hyperflexion mode is 25 N and 27.2 N, correspondingly. The average maximum displacement is 50.6 mm for the middle finger and 51.3 mm for the little one. Although the average maximum load and the average maximum displacement values are similar for both fingers, the dispersion among the obtained values is bigger for the case of the middle finger. Figure 13a and Figure 13b present the load-displacement plots for the middle and little fingers obtained from the flexion tests in abduction condition. The average maximum load for the middle finger in hyperextension and abduction condition is 25.0 N and 29.6 N, correspondingly. For the little finger, the average maximum force is 27.2 N in hyperextension mode, and 47.9 N in abduction condition. The little finger can withstand bigger loads in both test conditions due to the fact that the length in cantilever mode is shorter, compared to that for the middle finger. The test curves exhibit discontinuities due to re-accommodation of the finger assembly. Table 4 presents the set of tested samples and the corresponding designation.



(a) (b)

Figure 12. Finger hyperextension test curves for the middle (a) and little (b) fingers of the D1M hand. The horizontal dashed lines represent the average of the maximum load values.



(a) (b)

Figure 13. Finger abduction test curves for the middle (a) and little (b) fingers of the D1M hand. The horizontal dashed lines represent the average of the maximum load values.

As shown in Figure 14a, failure of the fingers in hyperextension occurs at the link of the proximal phalanx. On the other hand, failure in the abduction condition takes place at the MCP joints, as observed in Figure 14b; cracks nucleate in this area and propagate producing misalignment of the pins. This mechanism might be the reason why significant variations in the load-displacement curves beyond the linear region are evidenced (see Figure 15). Despite being thinner and thus having a smaller cross-sectional area for load distribution, the little finger can withstand a greater load than the middle finger due to the shorter moment arm from its base to the fingertip, which should be taken into consideration for the design.

In the case of the fatigue test, after 24000 cycles of loading, no component from the finger assembly (3D printed pieces, metallic pins, and wires) show any type of failure. Table 5 shows the weight of the fatigue/wear test samples before and after the test. A relatively small difference in weight was recorded as a result of the cyclic load and wear on the contacting parts, the average percentage of variation of weight was 0.05%. The 24000 cycles represent a period of 27 days of use of the prosthetic hand, considering that an average person performs 3680 grasp movements per day [22] and that a prosthetic hand user employs it 4 hours per day [23]. Based on the obtained results, the expected annual weight loss of the D1M hand would be 0.67%, which is a low value when compared to the wear rate for common materials. Figure 16 shows one of the fatigue test samples after the test; the encircled area highlights the ABS material particles resultant from the wear on the contacting parts.

Table 4: Designation of test specimens.

Test type	Sample	Number of samples	Samples designation
Hyperflexion	Middle finger	3	Hm1, Hm2 and Hm3
Hyperflexion	Little finger	3	Hl1, Hl2 and Hl3
Abduction	Middle finger	3	Am1, Am2 and Am3
Abduction	Little finger	3	Al1, Al2 and Al3

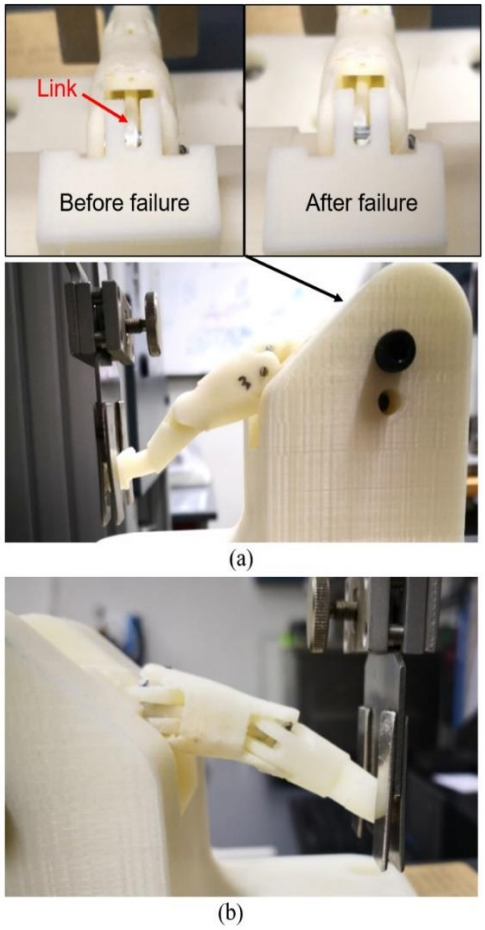


Figure 14: D1M prosthetic hand finger failure location under hyperextension (a) and abduction (b) conditions.

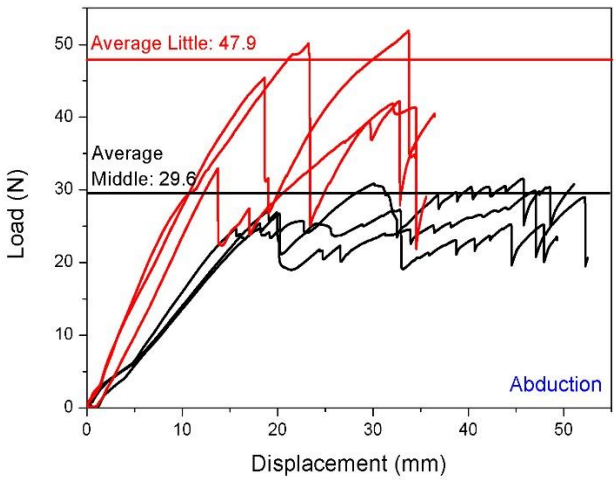


Figure 15: Finger abduction test curves for the middle (black) and little (red) fingers of the D1M hand. The horizontal lines represent the average of the maximum values per finger.

Table 5. Weight of fatigue/wear test samples before and after the test.

Sample	Weight (g)		Weight loss (%)
	Before test	After test	
1	6.7117	6.7109	0.012
2	6.7090	6.7020	0.104
3	6.8066	6.8040	0.038

Figure 16: Fatigue test sample

4. Conclusions

In this work, mechanical testing methods for upper limb prostheses were adapted and applied to a specific design, the D1M hand. Details of the experimental procedure have been described and test results have been discussed. The maximum tensile load that the D1M prosthetic hand can withstand is 112.4 N. Furthermore, maximum flexion strength of fingers in hyperextension and abduction conditions is 27.2 N and 48 n, respectively. It was also determined that after 24000 cycles of flexion no component of the fingers assembly of the D1M prosthetic hand exhibits any type of failure.

Obtained mechanical resistance tests results can be utilized to compare future experimental and simulation results. Moreover, our findings and the case described contribute in generating data for the 3D-printed prostheses community since standardization is needed in order to evaluate properties and performance of existing and future designs, which will allow for objective comparison between them and the continuous improvement of their mechanical specifications.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work was funded by the National Council of Science, Technology and Innovation of Peru - Cienciactiva from Concytec (Grant Number 163-2017). The authors would like to thank Carlos Romero and Jorge Pajuelo for their support in the testing machines' operation.

References

[1] B. Phillips, G. Zingalis, S. Ritter and K. Mehta, "A review of current upper-limb prostheses for resource-constrained settings," in IEEE Global Humanitarian Technology Conference (GHTC), Seattle, WA, USA, 2015.

[2] A. E. Tontowi, L. Ramdani, R. V. Erdizon and D. K. Baroroh, "Optimization of 3D-Printer Process Parameters for Improving Quality of Polylactic Acid Printed Part," International Journal of Engineering and Technology (IJET), vol. 9, no. 2, pp. 589-600, 2017.

[3] J. Torres, M. Cole, A. Owji, Z. DeMastry and A. P. Gordon, "An Approach for mechanical property optimization of fused deposition modeling with polylactic acid via design of experiments," Rapid Prototyping Journal, vol. 22, no. 2, pp. 387-404, 2016.

[4] B. Tymrak, M. Kreiger and J. Pearce, "Mechanical properties of components fabricated with open-source 3-D printers under realistic environmental conditions," Materials and Design, vol. 58, pp. 242-246, 2014.

[5] M. Fernandez-Vicente, W. Calle, S. Ferrandiz and A. Conejero, "Effect of Infill Parameters on Tensile Mechanical Behavior in Desktop 3D Printing," 3D Printing and Additive Manufacturing, vol. 3, no. 3, pp. 183-192, 2016.

[6] A. R. Torrado and D. A. Roberson, "Failure Analysis and Anisotropy Evaluation of 3D-Printed Tensile Test Specimens of Different Geometries and Print Raster Patterns," Journal of Failure Analysis and Prevention, vol. 16, no. 1, pp. 154-164, 2016.

[7] J. Cantrell, S. Rohde, D. Damiani, R. Gurnani, L. DiSandro, J. Anton, A. Young, A. Jerez, D. Steinbach, C. Kroese and P. Ifju, "Experimental Characterization of the Mechanical Properties of 3D Printed ABS and Polycarbonate Parts," Rapid Prototyping Journal, vol. 23, no. 4, pp. 89-105, 2017.

[8] R. Mio, M. Sánchez and Q. Valverde, "Mechanical Testing Methods for Body-Powered Upper-Limb Prostheses: A Review," in IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE), Taichung, 2018.

[9] S. Tenim, Design of an Affordable Anthropomorphic Mechanical Prosthetic Hand, M. A. Thesis, University of Cape Town, South Africa, 2014.

[10] J. Koprnicky, J. Safka and M. Ackermann, "Using of 3D Printing Technology in Low Cost Prosthetics," Materials Science Forum, vol. 919, pp. 199-206, 2018.

[11] Robotiq, "How we tested the Robotiq Adaptive Gripper," [Online]. Available: https://blog.robotiq.com/bid/31435/How-we-Tested-the-Robotiq-Adaptive-Gripper. [Accessed 7 Junio 2018].

[12] J. J. Crisco and S. W. Wolfe, "Mechanical testing of orthopedic implants: Hand and wrist," Mechanical Testing of Orthopaedic Implants, Vols. Duxford, United Kingdom, pp. 63-98, Woodhead Publishing, 2017.

[13] American Society for Testing and Materials, ASTM F1781 - 15, Elastomeric Flexible Hinge Finger Total Joint Implants, West Conshohocken: PA: ASTM International, 2015.

[14] T. J. Joyce and A. Unsworth, "A test procedure for artificial finger joints," Proceedings of the Institution of Mechanical Engineers,Part H: Journal of Engineering in Medicine, vol. 216, no. 2, pp. 105-110, 2002.

[15] International Organization for Standardization, ISO 22523 External limb prostheses and external orthoses, Switzerland: Requirements and test methods, 2006.

[16] National Institute of Standards and Technology, "Performance Metrics and Benchmarks to Advance the State of Robotic Grasping," [Online]. Available: https://www.nist.gov/programs-projects/performance-metricsmetricsand-. [Accessed 4 June 2018].

[17] J. T. Belter, M. T. Leddy, K. D. Gemmell Jr. and A. M. Dollar, "Comparative Clinical Evaluation of the Yale Multigrasp Hand," in IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob), UTown, Singapore, 2016.

[18] R. U. N. &. Media, "A hand for the hand-makers," 3 April 2016. [Online]. Available: http://news.rice.edu/2016/04/13/a-hand-for-the-hand-makers-3/. [Accessed April26 2018].

[19] M. Bustamante, R. Vega-Centeno, M. Sánchez and R. Mio, "A Parametric 3D-Printed Body-Powered Hand Prosthesis Based on the Four-Bar Linkage Mechanism," in IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE), Taichung, 2018.

[20] G. Smit, D. H. Plettenburg and F. C. T. van der Helm, "The Lightweight Delft Cylinder Hand: First Multi-Articulating Hand That Meets the Basic User Requirements," in 431-440, 23, 2015.

[21] J. T. Belter, J. L. Segil, A. M. Dollar and R. F. Weir, "Mechanical design and performance specifications of anthropomorphic prosthetic hands: A review," Journal of Rehabilitation Research & Development, vol. 50, no. 5, pp. 599-618, 2013.

[22] V. Nagaraja, J. Bergmann, D. Sen and M. Thompson, "Examining the needs of affordable upper limb prosthetic users in India: A questionnaire-based survey," Technology and Disability, vol. 1, pp. 1-10, 2016.

[23] M. Vergara, J. Sancho, V. Gracia and A. Pérez, "An introductory study of common grasps used by adults during performance of activities of daily living," Journal of Hand Therapy, vol. 27, pp. 225-234, 2014.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 69-81 (2019)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Assessment of the Quality and Sustainability Implications of FIFO and LIFO Inventory Policies through System Dynamics

Phoebe Lim Ching[*, 1, 2], Jose Edgar Mutuc[1], John Anthony Jose[3]

[1]*Department of Industrial Engineering, De La Salle University, 1004, Manila, Philippines*

[2]*Department of Industrial Engineering and Decision Analytics, Hong Kong University of Science and Technology, Kowloon, Hong Kong*

[3]*Department of Electronics and Communications Engineering, De La Salle University, 1004, Manila, Philippines*

ARTICLE INFO

Article history:
Received: 20 June, 2019
Accepted: 11 August, 2019
Online: 03 September, 2019

Keywords:
System dynamics
Inventory management
FIFO

ABSTRACT

Perishable inventory management contributes simultaneously to society and the economy, by reducing food wastage and capitalizing on the freshness of goods. For this reason, countless mathematical models have been developed for their effectiveness and cost-efficient management. Yet, the majority of these models can only optimize systems for a limited time frame, allowing for small gains in operations management, but failing to change the recurring patterns in inventory levels. System dynamics (SD) modelling shifts emphasizes these patterns and the recurring decisions that make them. Moreover, the framework has generated insight for other supply chain cases that could not have been derived from a short-term perspective. Thus, the current study now seeks to apply the SD framework in modelling perishable inventory systems, in designing policies that benefit the environment and the economy by reducing waste production and increasing the viability of goods reaching the customer. In particular, it evaluates the impact of opposing issuance policies (i.e. First-In-First-Out (FIFO) and Last-In-First-Out (LIFO)) on perishables to demonstrate the potential of SD in improving perishable inventory management. The simulated results share the sentiments of optimization models, that FIFO will ultimately generate less wastes and incur less material costs. Yet, the simulations also reveal implementing FIFO will result in larger fluctuations in inventory levels, which imply greater inconsistency in age-based quality. These suggest that LIFO would be preferable for quality-sensitive products, while FIFO would be preferable for cases sensitive to waste production. The current study demonstrates the efficiency of system dynamics in generating insight beyond that which can be derived from the existing mathematical models. Future studies may likewise extend this approach in the evaluation of policies on the use of technology in perishable inventory systems, which are the prevalent in present literature.

1. Introduction

In inventory models, perishability is a physical property of inventories that results in the eventual expulsion of these goods from the system as waste [1]–[3]. In terms of social and commercial impact, studies on the management of perishables have prevailing importance. Products necessary to sustain human life, such as food and healthcare, possess this characteristic [4], [5]. In particular, the case of blood banks has received much attention from studies on perishable inventories. Due to the instability of supply and short validity of blood units, a precise and quantified means of managing these inventories are much needed [4], [6]. Blood banks are a strong example of how perishability as a product characteristic can complicate inventory systems, and can be used to generate insight on how other perishable products are best to be managed [6].

Perishability complicates the traditional inventory and supply chain models. The typical components of such models are profit and sustainability measures, capacity constraints, and demand

[*]Corresponding Author Phoebe Ching, phoebe_ching@dlsu.edu.ph

distributions [7], [8]. However, policies that prove optimal for non-perishables are not necessarily optimal for perishables [5], [6]. In some models, perishability is represented as the deterioration of product quality [5], [9]. This can affect demand or the allowable pricing, thereby sharply differentiating the scenario described from inventory systems that do not differentiate between fresh and aged inventories. The transition of inventories from a state of salability to wastes is another differentiating factor. The production of wastes affects system capacity [10], [11], and any wastes generated may incur disposal costs [11], [12]. In this case, perishability affects key system constraints and parameters. Some studies on this subject have been dedicated to studying the differences between perishable and non-perishable systems. These have proven that the differences between perishables and non-perishables result in different optimal policies for each kind of inventory.

The majority of studies in this area are optimization models. From the models themselves, decisions on production quantity and product timing can be determined for a given set of parameters. In turn, managerial insights on managing inventory given similar parameters can be derived from the quantitative behavior of the variables. The optimization of inventory systems is a continuously growing body of research, with models growing in scale and complexity, or through the incorporation of new components such as technology enhancements [3], [13]. Aside from this, research efforts are also being invested in identifying the most efficient algorithms for the model [9], [14]. All in all, advancements in this area are geared towards achieving greater accuracy and numerical exactness. However, a weakness of this approach could be that the emphasis on these aspects of modelling shifts the focus away from insight-generation. While the models are becoming increasingly adept at describing specific scenarios, it may be noted that real inventory systems are constantly subject to external change. Thus, regardless of the level of exactness that is achieved by these models, its direct applicability to real systems will always be limited.

In response to these issues with current modeling approaches, applying the system dynamics methodology can augment these through (1) its applicability to large-scale systems which mathematical modeling may be inefficient for [15], (2) its emphasis on patterns of behavior over numerical exactness [16], [17], and (3) its limitation to constant system elements, as opposed to designing in response to external factors [18], [19]. The first two attributes make system dynamics apt for identifying key areas for solution implementation. Specifically, it can be used to identify the most influential variables in the system. This is insightful for future studies, which can delve more specifically in those areas. The third attribute aids in generating insight for recurrent managerial policies as opposed to period-specific decisions. These are more aligned with real managerial practices which, where rules and guidelines are favored to specific values.

In this study, the system dynamics (SD) methodology is applied to the case of perishable inventory systems. This is done for the purpose of demonstrating how the SD methodology can be applied in designing managerial guidelines, with proof that use of these guidelines will generate significant improvement for the system. As SD has already generated insights for general supply chain settings [20], [21], this study now investigates how perishability can alter the results, given that perishability has been proven to be a significant differentiating factor in previous studies [22]–[24]. Through its high-level perspective, the current study aims to identify impactful areas for solution development in the perishable inventory scenario. In this way, it can contribute to the area of study with suggestions for future studies focusing on these specific areas.

2. Literature Review

2.1. Inventory System Dynamics

There seems to be a limited research concerning inventory system dynamic (ISD) studies, thus there is potential to address multiple unexplored research gaps [19]. The challenge then is to find a purposeful gap to address, and justifying why system dynamics (SD) methodology would be more appropriate over methods that allow for more quantitative accuracy and optimization. In this respect, a review of the existing ISD studies is conducted to (1) learn the research objectives that prompted the use of SD, (2) understand how the ISD study relates to other inventory models using different methods, (3) understand how managerial and policy-making insights are derived from such studies, and (4) determine the main contribution of individual ISD studies.

In a literature review on SD applications in renewable energy supply chains, Saavedra 2018 observed that related studies seems to have objectives of either (1) improving understanding of the supply chain; or (2) developing modelling and simulation approaches. As a framework for understanding, SD is effective in depicting complex cause-and-effect chains and various sectors through its diagram-based modelling approach. From its diagramming methods alone, insight can be derived from the network of interrelationships between variables. On the other hand, as a simulation tool, the effectiveness of SD in providing insight is still being validated. This is due to its lack of quantitative accuracy and the lack of studies in this area.

Generally, a well-accepted method of applying SD is to use the diagramming approach to identify problematic areas, and apply sensitivity analysis to determine the most efficient area for solution [25]. This is the approach taken by studies seeking to apply the SD framework in a specific context. On the other hand, in studies that assess the methodology itself, SD has been coupled with various methods to augment its lack of numerical soundness, and the lack of perspective in other modeling approaches [19].

Given its frequent applications in sustainable supply chain models, there is often a part of SD studies that compares its methods with the likes of operations research and life cycle assessment (LCA), and validates that it can produce insights beyond those that can be derived from other modeling approaches. In comparing supply chain models using SD with those using other methodologies, Saavedra 2018 concluded that a major difference between SD and LCA is its focus on the behavioral patterns of variables within a model, such as the intermittency and variability of its values. On the other hand, the focus on LCA is on aggregate performance, such as the total volume of emissions [18].

The SD methodology also allows for a wide range of variables to be modeled and simulated efficiently, given its lack of numerical

exactness. In a literature review, it was observed that most SD supply chain models had a macroscopic perspective consisting of economic and environmental impact. Next in quantity were SD studies with an inter-organizational scope, and last were studies with an intra-organizational scope. On the contrary, for operations research-based studies, the order from highest to lowest volume of publications is intra-organizational, inter-organizational and macroscopic [19]. This can alter the kinds of policies that may be developed and simulated using each method.

SD can be a means of extending optimized parameters into concrete policies, particularly in cases with multiple parameters of a different nature. Song 2018 created a model to simulate the application of optimized parameters for technology, environment, energy and economy in a system where they are interrelated. These parameters were derived using data envelopment analysis (DEA), which augments the lack of numerical accuracy in most SD models. SD then strengthens the policy-making approach by highlighting the physical relationship between variables.

Scenario analysis, such as applying different parameters to represent different cases, is a quick means of validating the model's logic and policy development. In validation, it allows for multiple real-life case parameters to be applied to prove the validity of the causal relationships and the completeness of the model. In policy-making, a number of alternative solutions can also be generated and simulated, by shifting model parameters [18].

Overall, SD studies have contributed to the field of study with insight for how certain resources can be managed with optimal results [25]; knowledge on the opportunities and limitations of simulation methods [18]; and the creation of hybrid models combining SD with more math-based methodologies [19].

2.2. Perishable Inventory Systems

Perishable Inventory Systems (PIS) have received much attention in literature as early as the 1990s, up until recent time [5]. Earlier studies simply sought to incorporate 'perishability' as an attribute of inventory. Nonetheless, more recent studies seek to determine the kinds of managerial strategies and policies that may be applied to perishable inventories [2], [26]

In a literature review conducted in [5], studies from 2012 to 2015 were segregated according to the different policies being implemented in each study. It was found that the top policies being studied during that period were (1) credit and payment policies, (2) supply chain information sharing, and (3) pricing and markdown problems. There were notably few studies on the use of the introduction of advanced monitoring technology, and on classic issuance policies (i.e. FIFO, LIFO).

Since then, there has been an increase in models on the integration and use of monitoring technology in perishable inventory systems. [6] have introduced the use of RFID technology in monitoring of unit age and cost to the dual-sourcing problem for perishables. This was evaluated against fixed and exponential product lifetime scenarios, to gain insight on specific potential applications for this strategy. More recently, [3] have developed an IoT-based model for monitoring perishables in shared storage. Given the potential for cross-perishability, which refers to the increase or decrease in deterioration rate resulting from shared storage, information monitoring in this manner could then allow for lowering of deterioration costs, inventory level, and quality loss for the system. On a two-stage optimization model for pricing and replenishment of PIS, the recommended extension is also the incorporation of RFID and point-of-sales data in decision-making [7].

It seems to be reasonable for research to progress in this direction given the rise of Industry 4.0 [27], [28]. However, the implementation of advanced systems, similar to the aforementioned, are still dependent on fundamental decisions in inventory management. This is evidenced by the way that these technologies are modeled in conjunction with inventory management policies from classic literature [29]. Yet, these policies are still among the top policies being studied from 2012 to 2015, as identified by [5].

Over the years, there have been few changes to the performance functions used for perishable inventory models which enable ease of comparison. Financial measures such as profit and cost are still used in almost every PIS study. The percentage of profit growth or cost reductions resulting from the implementation of certain policies is typically used as the measure of performance in these cases [4], [6], [26]. Profit can be affected by perishability through the decline of demand alongside quality [30]. It can also manifest as disposal costs [2], [8] and wastes [7].

Wastes differ from financial measures as they are a physical outcome of the PIS policy. By quantifying waste production, a model can also be assessed for its sustainability [10], [11]. The minimization of waste, separate from its contributions to cost, is used as an objective in a number of studies [5]. In this way, the model can differentiate between financial and physical outcomes, given that the volume of wastes would not always be economically relevant in all cases.

Another physical outcome of PIS policy is the service level, or the ability to meet customer demand [28]. In the model of [7], service level is the probability of stockout in the replenishment cycle resulting from the policy. Monitoring the service level of PIS prevents over-emphasis on cost efficiency, in a more tangible way than revenues. Aside from this, the model of [4] has made use of the average age of inventories being sold, to represent the quality of the products.

The perishable nature of inventories in PIS is represented as a finite lifetime. This leads to wastes at the end of the product lifetime [11], [12], and declining demand over time for a specific unit [5], [9], [31]. In most PIS models, a scenario analysis is conducted for various rates of deterioration. Generally, the higher the deterioration rate, the greater the amount of waste [7]. However, this can also vary depending on the stochastic distribution of a product's lifetime, and the kind of policy being implemented. The model of [6] demonstrates that dual-sourcing can produce greater cost savings in cases where lifetime is fixed with a shorter average, and when lifetime is exponentially distributed with a longer average.

Generally, a number of policies have proven effective in managing PIS, with immediate improvements in cost and material efficiency. Optimization models demonstrate how such policies can be applied with optimal results for a given set of parameters.

However, the existing methodologies have proven to be inefficient for policy-to-policy comparison, given the variability in model structures and sensitivity to parameters. Thus, the basis of determining best practices based on the existing studies could still be improved.

2.3. FIFO and LIFO Issuance Policies

Issuance policies are an integral part of an inventory management problem. In general, the parameters under study in such a problem are pricing, inventory costs, and deterioration rate [7], [32]. These parameters are present in almost all studies and are used to evaluate specific policies being applied, or intended for application, on real inventory systems. These policies can include issuance policies (i.e. FIFO and LIFO), which relate to the order at which goods of the same kind at different ages are dispatched to the customer. Based on the review of Janssen (2016), issuance policies appear to be an overlooked area in literature. However, based on the current review that was conducted, it seems that most studies include issuance policies as a parameter, instead of treating it as a focus of the study. The review of literature in this area is thus segmented in two ways: (1) studies considering issuance policies as a parameter, and (2) studies explicitly focused on evaluating or identifying optimal issuance policies.

Using a policy as a parameter means that it can be changed to allow the analyst to see various scenarios. [23] sought to determine how the centralization of inventory could be beneficial in managing a perishable supply network. Specifically, the case under consideration was a blood supply chain with inventory located at a single blood bank and being supplied to multiple hospitals. FIFO was explicitly modeled for this case, as this is the standard for inventories in the healthcare service industry. [10] adapted a modeling and simulation analysis methodology determine the impact of a closing day constraint for the inventory of stores. The objective of the study was to determine if the existence of a closing day would lead to more aged inventory as closing day approach, as is the case in real grocery store inventories. It was found that the closing day constraint did result in simulated behavior being closer to the real behavior of inventories. In another study, [11] assessed the quantitative impact of micro-periodic inventory management policies, for how these may potentially improve costs and waste production. Under this approach, a single day was broken into morning, midday, afternoon and evening periods. This likewise made use of a mixed FIFO and LIFO approach. It was found that costs were reduced significantly from micro-periodic management, as compared to daily planning.

One reason why a mixed approach is taken in most studies is so that the exact impact of each policy will not have to be considered. This recognizes that there are, in fact, real differences between the effects of applying each policy, particularly on perishables. The models that explicitly analyze the issuance policies seek to determine what these effects are and how they can be managed most effectively to meet the goals of the system. [12] compared the viability of FIFO and LIFO for managing perishable inventories using a modified Economic Manufacturing Quantity. Said model was modified to give consideration for fixed order quantity and joint ordering policies. This revealed that LIFO could prove optimal under linearly decreasing price structures. This finding is relevant given that it is generally accepted that FIFO has superior performance on perishables [19], [23].

In most cases, issuance policies are modeled as a two-warehouse problem. Under LIFO, the first inventories to be sold would come from the rented warehouse; while under FIFO, the first sales are taken from the owned warehouse. A model for this problem was developed by [32], who assessed each issuance policy for its overall costs, given deterioration and holding costs. It was found that FIFO was less expensive when the cost parameters were lower for the owned warehouse than in the rented warehouse. As this is the case for most entities, it may be concluded that the study recommends FIFO under LIFO given the parameters that were considered. [9] extends the study of issuance policies for two-warehouse inventory systems to the use of modified FIFO (MFIFO) and modified LIFO (MLIFO) policies. Under the MFIFO rule, the FIFO rule is followed as far as the inventory in demand is available. When unavailable, the backorder gets assigned to the lowest priority level in the next period [28]. The model was structured in such a way that allowed for easy comparison between policies by relaxing a few constraints. It was found that when deterioration rate was different between the rented and owned warehouses, holding cost could make a significant difference in the optimality of each policy.

MFIFO and MLIFO are already evolutions of the existing issuance policies. It is understood that FIFO and LIFO each have weaknesses, and thus improvements can be made in their exact rules. [31] proposed the development of a new issuance policy entitled Allowance-in-Fraction-Out which has specific rules for when to apply FIFO and LIFO. This is likewise applied to the two-warehouse inventory problem, wherein the decision to switch is based on the fraction of rented warehouse inventory to owned warehouse inventory. Across various cases of deteriorating behavior, value of information, and perishability, it was found that the mixed method could have benefits in transportation costs and the occurrence of sub-replenishments.

In recent literature, most studies treat issuance policies as a parameter. Moreover, most studies assume a mixed issuance policy structure to protect the model from the effects of using any specific policy. However, this suggests that inventory will be taken randomly from a company, which is understandably not the case in inventory systems. The most advanced study in this area seeks to determine the appropriate rules for using FIFO and LIFO, and managing their negative repercussions. This would imply that further study is needed of the explicit behaviors of FIFO and LIFO to determine how they may be managed productively.

3. Methodology

In this study, the system dynamics (SD) methodology is used to model a perishable inventory model, and simulate the application of certain policies and methods. This method is efficient for analyzing patterns of behavior, as it emphasizes long-term patterns over quantitative accuracy [16], [17]. It is also efficient in expanding the scope of a model to encompass all variables that could have significant feedbacks on the key variables under study [20]. SD emphasizes the importance of feedbacks in a system, showing how the system's behavior is actually the result of these internal feedbacks, and not external factors [16]. In problem solving, this maintains the focus on constant factors,

rather than externalities which may or may not be present. In this way, the SD methodology is well-suited for evaluating policies in the long-term, and managing any negative repercussions that may result from their application.

The SD framework is applied with the intention of designing policies that will improve the sustainability and quality goals of the system. Most perishable inventory models attempt to address sustainability goals by having waste minimization as one of its objectives [30], [33]. The connection of waste minimization to sustainability implies that through waste reduction, resource consumption is also reduced. On the other hand, quality goals can be met in terms of age-based quality, by minimizing the time a unit spends in the system from production to sale [4]. Age or, conversely, freshness is an apt representation of quality for perishables given the fact that (1) perishables are typically fast-moving consumer goods which vary little from unit to unit, and (2) perishables will definitely degrade in some way over time, making age a significant component of quality, even if it is not the sole component.

The SD methodology classifies variables as stock, flow, auxiliary or exogenous based on their role in the model. Stock variables accumulate over time based on previous values; flow variables represent this rate of accumulation; auxiliaries serve as the transition of stocks to flows; and exogenous variables are parameters whose values are externally determined. A discussion of how the variables were classified, and of their respective equations is given in Section 3.1. Following this, the variables determined in Section 3.1 are organized in a causal loop diagram in Section 3.2. This represents the relationship of the variables in an enclosed system. The variables and loop structure are then integrated in a stock-flow model for simulation and analysis in Section 3.3.

3.1. Variables Definition

Stock variables are those which increase or decrease in relation to their value in a previous period. This can be used to represent inventory levels which physically change over time. In this model, inventory is divided into two categories based on age. Hence, there is a variable representing Fresh Finished Goods Inventory (FFGI) and Previous Finished Goods Inventory (PFGI) respectively, as levels. By logic, FFGI is a function of production, sales, and leftover inventory in the current period (t). Since it is not an option to dispose freshly-produced goods in this model, there is no difference between the equation for FFGI in the perishables and non-perishables case.

$$\text{FFGI}(t) = \text{FFGI}(t\text{-}1) + \text{Prod.}(t) - \text{Sold F. Inv.}(t) - \text{Held Inv.}(t) \quad (1)$$

PFGI represents inventory carried over from previous periods. In the perishable's scenario, a fraction of inventory becomes waste at the end of each period. Mathematically, these differences between the perishables and non-perishables case are given in equations 2 and 3 respectively.

$$\text{PFGI}(t) = \text{PFGI}(t\text{-}1) + \text{Held Inv}(t) - \text{Sold P. Inv}(t) - \text{Wastes}(t) \quad (2)$$

$$\text{PFGI}(t) = \text{PFGI}(t\text{-}1) + \text{Held Inv}(t) - \text{Sold P. Inv}(t) \quad (3)$$

Aside from variables that physically change in value, stock variables are also apt for representing human cognition in system dynamics models. Cognition is influenced by the occurrence or lack thereof of certain events. In consequence, these strengthen or weaken a certain idea, which in turn produces reactions affecting the other variables in the system. Such variables are usually not included in modeling as there is no definite metric to represent them. However, the soft emphasis on numerical accuracy in system dynamics allows for such variables to bear the same weight as tangible variables (i.e. inventory, cashflow).

In this model, there are two cognition variables, both represented as 'perception'. One of these is Perceived Quality (PQ), representing the customer's flawed knowledge on the actual quality of goods produced by the producer. System dynamics recognizes the feedback delay in perception of performance, which results in there appearing to be no effect despite efforts toward improvement. PQ changes by a positive or negative value, depending on whether the quality in recent periods has been good or bad. However, PQ itself will always be a normalized value as the change happens in relation to previous values of PQ. Conversely, Perceived Demand (PD) represents the producer's flawed knowledge on the actual volume of customer demand. In real applications, this would represent the degree of error in demand forecasts [34]. Likewise with PQ, PD changes by a positive or negative value depending on the recent demand values.

The change in the value of stock variables at each period is defined as a flow variable. A stock can have inflows that add to its value, and outflows that reduce its value. In the case of FFGI, the inflow rate is Production (Prod), while the outflow rates are Held Inventory (Held Inv) and Sold Fresh Inventory (Sold F. Inv). Prod is based on the value of PD, which represents the use of forecasts in deciding on production quantities [32], [35].

$$\text{Prod}(t) = \text{PD}(t) \quad (4)$$

Held Inv is equal to the current value of FFGI, as all fresh inventory that is not sold will be carried over to the next period as PFGI. Held Inv represents the transition of FFGI into PFGI by being an outflow of the former into the latter. The also represents the aging of products in the system. In essence, there are only two age classes of goods--fresh and previous. While this may not capture the exact age of the inventory pieces, it is nonetheless sufficient to show the impact of FIFO and LIFO. Further segmentation by age group may further complicate the model without generating additional insight to compensate for the effort.

$$\text{Held Inv}(t) = \text{FFGI}(t) \quad (5)$$

FFGI and PFGI each have outflows in the form of sales, and the equations for these vary between the FIFO and LIFO cases given that the former prioritizes selling PFGI while the latter prioritizes selling FFGI [Parlar et al, 2011]. These are Sold Fresh Inventory (Sold F. Inv) and Sold Previous Inventory (Sold P. Inv). Mathematically, the prioritized inventory class is directly influenced by Customer Demand (Cust Dem), while the deprioritized class is a function of the sales for the prioritized class. The equations for the FIFO case are given in equations 6 and 7; and those for LIFO are given in equations 8 and 9.

$$\text{Sold F. Inv}(t) = \text{MIN}(\text{MAX}(\text{Cust Dem}(t) - \text{Sold P. Inv}(t), 0), \text{FFGI}(t)) \quad (6)$$

$$\text{Sold P. Inv}(t) = \text{MIN}(\text{Cust Dem}(t), \text{PFGI}(t)) \quad (7)$$

$$\text{Sold F. Inv}(t) = \text{MIN}(\text{Cust Dem}(t), \text{FFGI}(t)) \quad (8)$$

$$\text{Sold P. Inv}(t) = \text{MIN}(\text{MAX}(\text{Cust Dem}(t) - \text{Sold F. Inv}(t), 0), \text{PFGI}(t)) \quad (9)$$

For PFGI, there is an additional outflow in the form of Wastes. This variable is only present in the perishables case, as inventory can only be accumulated in the case of non-perishables. The variable Wastes is a function of the residual PFGI and a given rate of deterioration:

$$\text{Wastes}(t) = (\text{PFGI}(t) - \text{Sold P. Inv}(t)) / \text{DelayP} \quad (10)$$

For the cognition variables Perceived Demand and Perceived Quality, it was previously mentioned that the change could be positive or negative depending on the recent values of the perceived information. These changes are the flow variables Change in Perceived Demand (CIPD) and Change in Perceived Quality (CIPQ). Mathematically, a positive change would result if the current value is greater than the perceived value, and a negative change if otherwise. In effect, the current values are normalized against previous values. A feedback delay is used to represent the limited influence on current values against long-standing perception. In equation form:

$$\text{CIPD}(t) = (\text{QOGS}(t) - \text{PQ}(t)) / \text{DelayPQ} \quad (11)$$

$$\text{CIPQ}(t) = (\text{Cust Dem}(t) - \text{PD}(t)) / \text{DelayPD} \quad (12)$$

Aside from stock and flow variables, a system dynamics model contains auxiliary variables. These are components of flow variables. It can serve as a means of breaking down potentially-complex equations for flows, both for ease of understanding and to highlight the behavior of specific components. Auxiliaries are derived from the behaviors of stock variables, and can link these to flows in the model. Unlike stock variables they are not a function of their previous value. While they do not change the overall system behavior, they are nonetheless important for clarity in representation.

The model contains three auxiliary variables: Quality Multiplier (QM), Customer Demand (Cust Dem), and Quality of Goods Sold (QOGS). QM is an example of an auxiliary that exists for ease of understanding. Its value is equal to the value of Perceived Quality (PQ), yet the conversion is made to show the logic in making PQ part of the function of Cust Dem. Hence:

$$\text{QM}(t) = \text{PQ}(t) \quad (13)$$

Cust Dem is a function of two variables, QM and a randomized demand value that is exogenous to the system (Natural Range of Customer Demand, NRCD). In effect, the perception of good quality positively influences demand by a margin, while the perception of poor quality has an inverse effect. As indicated by its usage in the previous equations, customer demand is the actual amount of purchases made by the customer in a single period. This consequently affects the sales variables, Sold Fresh Inventory and Sold Previous Inventory. The equation for this variable is given as:

$$\text{Cust Dem} = \text{NRCD} * \text{QM}(t) \quad (14)$$

QOGS represents the quality of goods sold for a current period. In this model, the quality of the perishable product is solely based on freshness. Its value is derived from the amount of Sold Fresh Inventory (Sold F. Inv) in relation to Customer Demand (Cust Dem). Higher proportions of Sold F. Inv in Cust Dem would thus equate to better quality. As information on the quality of the producer's goods is transmitted to the customer through sales, QOGS has a direct but gradual influence on PQ [30], [36].

$$\text{QOGS}(t) = 1 - \text{Sold P. Inv}(t) / \text{Cust Dem}(t) \quad (15)$$

The variables that are not directly influenced by any variable in the model are exogenous variables. The belief that the system's behavior is primarily influenced by the feedbacks between its endogenous variables, rather than exogenous variables, are at the core of the system dynamics methodology. Hence, exogenous variables are typically represented by constant parameters, as proof that their value is of minor significance to the system's behavior.

Feedback delays are a common example of an exogenous variable. These represent delays in the transmission of information across the system. In equation form, these are variables that limit the flow of information to a cognition variable. In this model, these delays exist in the flow variables for Perceived Quality and Perceived Demand, as Delay in Perception of Quality (DelayPQ) and Delay in Perception of Demand (DelayPD). Physical delays are another common exogenous variable, representing rates of transition from one state to another. In this model, the physical delay is the Delay in Perishing (DelayP). This delays the transition of inventory into wastes, according to a supposed validity period.

An exogenous variable is also used to represent additional sectors of the system that can be simplified for ease of analysis. This can be done when the variables included in that sector are not explicitly being analyzed. In this model, the Customer Demand (Cust Dem) is understandably an outcome of Perceived Quality. Yet, it is also influenced by the availability of other options in the market; the demand of the customer's own customer; and numerous other factors depending on the specific case and scenario. In effect, Cust Dem will have a degree of variability due to these other variables. To represent this, a Natural Range of Customer Demand variable is introduced as a random variable with a given range.

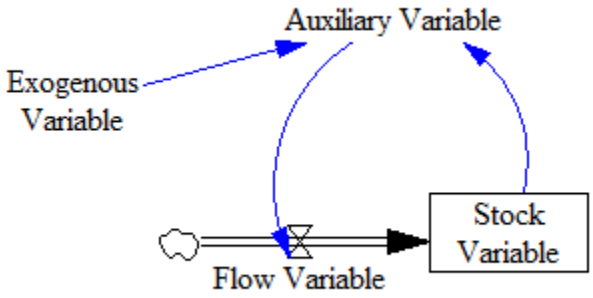


Figure 1: Basic Structure of a System Dynamics Model.

3.2. Interactions Between Endogenous Variables

The greatest differentiating factor of the SD methodology is that it organizes variables into feedback loops. This follows the principle that a system's behavior is influenced more by the system's internal reactions than any external factor. A visual representation of the role of each kind of variable in a feedback loop is given in Figure 1. All loops include at least one stock variable, which shows how the current outcomes of a system result from behavior in a previous period. This progression of cause and

effect leading from past to current behavior is demonstrated through a progression of auxiliary variables. It is shown in Figure 1 that auxiliary variables are the link from stock variables to flow variables, which mathematically means that the past values of a variable would change its next-period values. Finally, the exogenous variables are externally-determined and not influenced by any other variable in the system. Hence, there is no feedback leading to an exogenous variable.

To demonstrate the cause-and-effect relationships in a SD model, variables are arranged in causal loops without differentiation between variables. The purpose of this is to emphasize the feedbacks within the system prior to explicit modeling. In developing the causal loop diagrams for this study, it became apparent that there would be different causal loops for the FIFO and LIFO cases respectively, due to the difference in inventory type affected by sales. Specifically, demand directly impacts aged inventory in the FIFO case, and fresh inventory in the LIFO case. Aside from this, both cases have the same loops, representing a common PIS structure to enable comparison.

In PIS models, demand changes based on the age of inventory given the fact the that goods are deteriorating over time [5], [9]. In this model, a quality variable determined by age is positively linked to demand, meaning that demand will increase as quality improves. Quality is positively affected by the quantity of fresh inventory, and negatively affected by the quantity of aged inventory. This is how the reduction in demand caused by low-quality or aged goods is interpreted in the model. The resulting demand has a delayed impact on production, representing how demand from previous periods can determine production levels in the coming periods through forecasts. Production quantities then add to the level of fresh inventory, and fresh inventory becomes aged inventory, by natural progression [4], [11], [37].

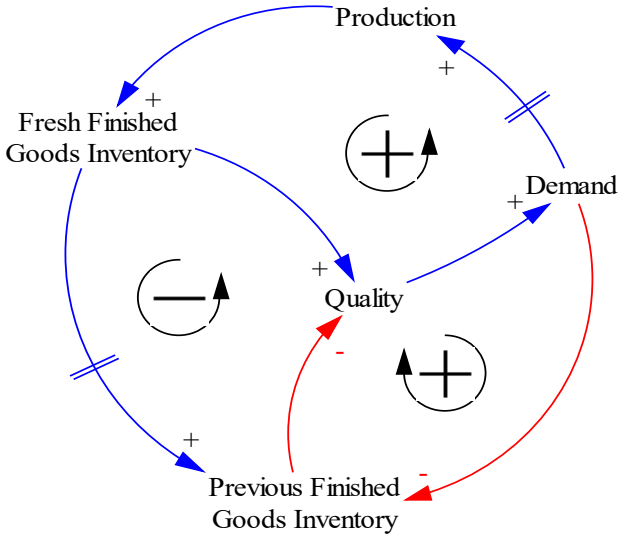


Figure 2: Causal Loop Diagram for FIFO Scenario.

The amount of positive or negative relationships in a feedback loop determine the 'positivity' or 'negativity' of that loop. Even numbers of negative relationships create a positive loop, which prompts positive exponential growth in the variables involved. Odd numbers create a negative loop with negative exponential growth. In system dynamics, systems are understood to be composed of these positive or negative loops, which balance each other out. Following this perspective, the LIFO scenario is more balanced than the FIFO scenario, as it contains equal amounts of positive or negative loops. This suggests that behavior is more consistent under the LIFO scenario, than the FIFO scenario.

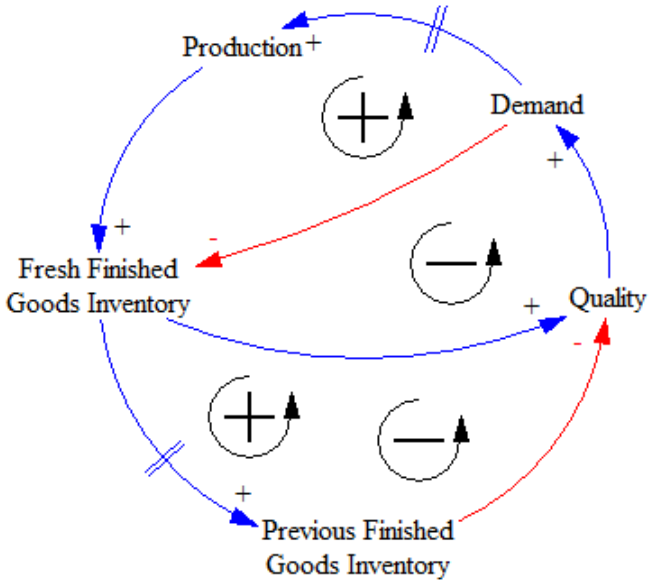


Figure 3: Causal Loop Diagram for LIFO Scenario.

3.3. Simulation Model Development

Using VensimPLE software, a system dynamics (SD) model was developed to simulate the postulated behaviors of each variable and analyze their interactions. The general form of the SD model is shown in Figure 4. This shows the variables identified in Section 3.1 in their proper form, and arranged in the feedback loops derived in Section 3.2. There are three major sections in the model: (1) the physical inventory system, which is also the main problem variable under study, (2) the perception of quality, representing the impact of the physical system on human receptors, and (3) the perception of demand, representing the human decisions which govern the physical system. These form a single loop that determines the behavior of the system.

The model parameters are feedback delays, which can be adjusted to represent various scenarios. The Delay in Perishing represents the validity period of the goods, which can vary depending on the nature of the goods. In PIS for blood banks, this parameter strongly impacts system behavior given the brief and stringent validity period of the inventory [23]. For disaster relief operations where goods are hardier, the impact of this parameter is less strong, given that the goods are practically non-perishable [2]. For the same reason, there are few disaster relief supply chain studies which consider waste and disposal.

The Delay in Perceived Quality represents how evident differences in quality may be to the customer. Smaller batch sizes and proximity to the end-customer can equate to shorter delays. In retail settings, which have this characteristic, demand is typically a function of quality [11]. On the other hand, upstream suppliers consider perishability in terms of waste and disposal costs, as opposed to impact on demand [23], [28].

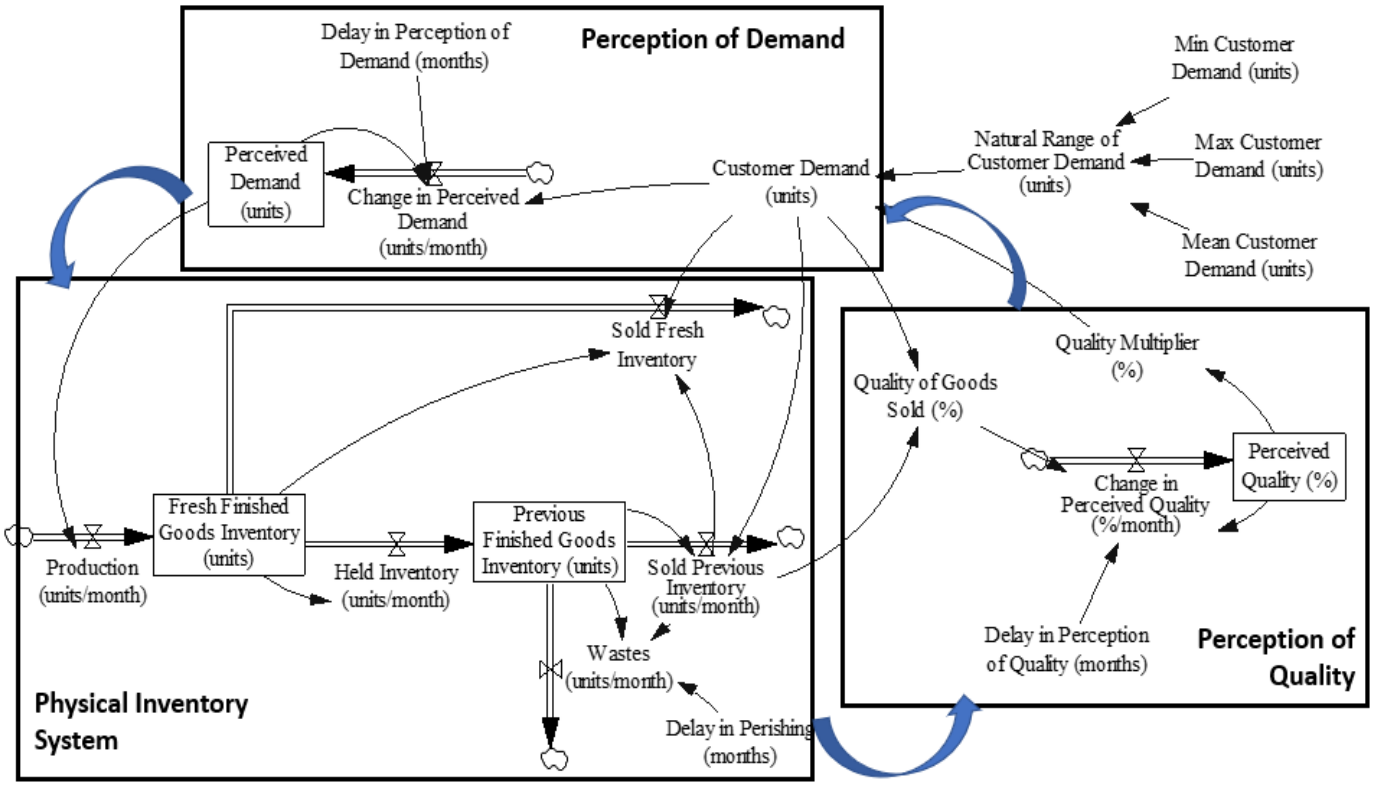


Figure 4: Structure of Inventory System.

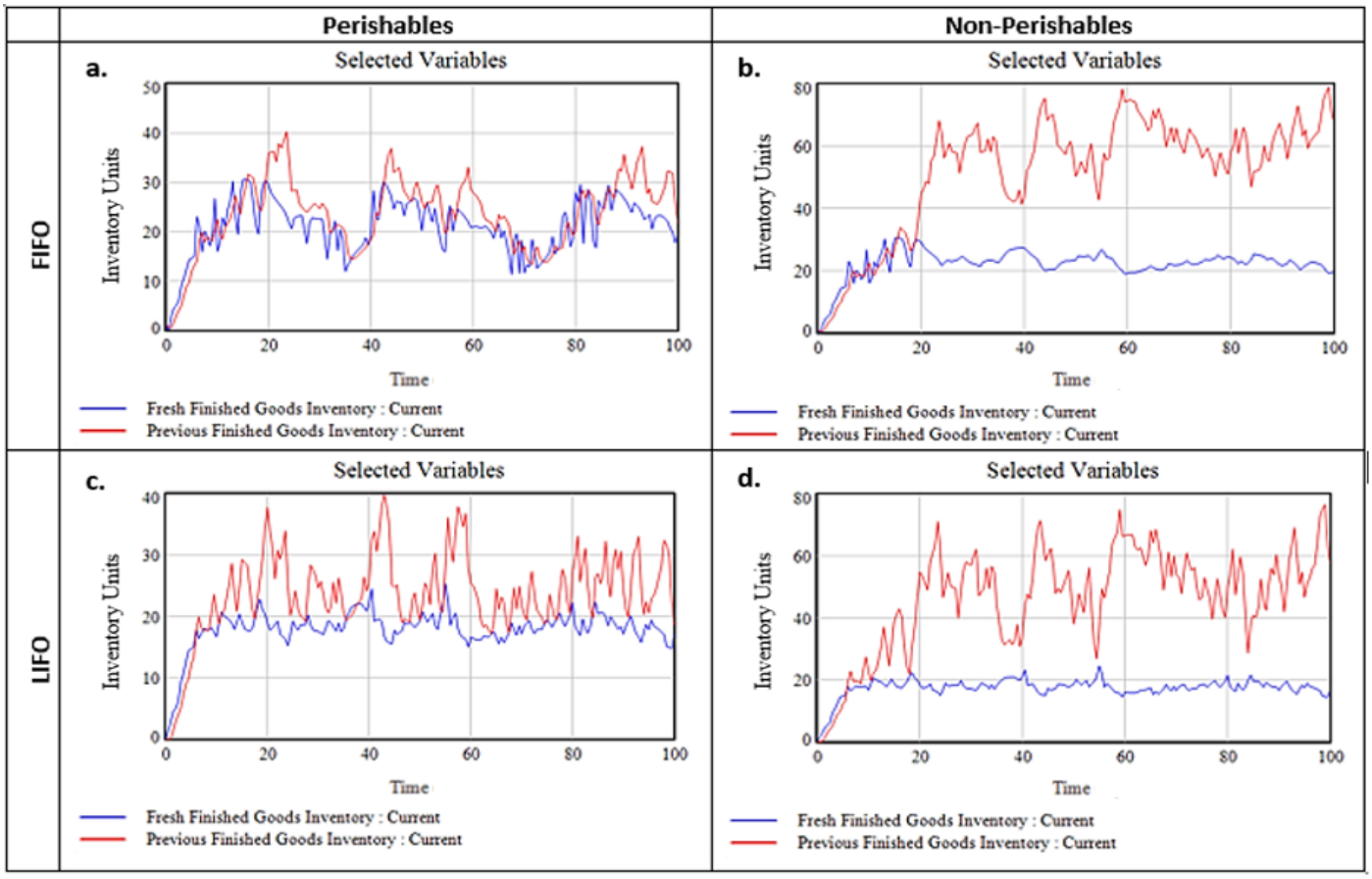


Figure 5: Simulated Fresh and Aged Inventory Levels.

The Delay in Perceived Demand represents inaccuracies in the forecasting techniques with regards to real demand. There are countless studies on forecasting techniques and replenishment policies, and each have proven effective in specific contexts and following specific demand distributions. In such cases, the policies are evaluated by the resulting service level [4], [7]. The current model extends this to examine the reactions between the performance of the policy and other components of the system.

Following the structure in Figure 4, four models were created for each of the following scenarios: (1) FIFO applied to perishables, (2) FIFO applied to non-perishables, (3) LIFO applied to perishables, and (4) LIFO applied to non-perishables. The main difference between the perishables and non-perishables scenarios is the lack of a Waste outflow for aged non-perishable inventories. In this case, the only possible outflow is through sales. Given that the emphasis of this study is on perishable inventories, the policies that may trigger the disposal of non-perishables are not considered.

By logic, FIFO and LIFO differ based on the priority of inventories to be sold. These differences were given in equation form in Section 3.1 and in diagram form in Section 3.2. It is necessary to generate a different model to represent each case given that the feedback relationships differ. Hence, this is not a case of merely adjusting parameters to represent a case, but making adjustments to the model parameter. Aside from priority of sales, the structure of the model given in Figure 4 is maintained for all versions, to allow for the results to be comparable with one another. The simulation results depict values for the behavior of variables over time. This is analyzed in graphical form to assess the model for patterns and identify areas for problem-solving.

4. Simulation Results

All inventory models have the objective of minimizing inventory levels, whether directly or indirectly. In optimization models, inventory levels are often linked to cost objectives through holding costs and disposal costs [15], [35]. In this way, minimal inventory levels becomes the most economic scenario for an inventory system. In simulation methodologies, inventory is physically monitored with the unstated understanding that it affects other performance indicators, such as cost and wastage [5], [7]. The system dynamics methodology is aligned with this perspective. This study likewise monitors physical inventory levels as a performance measure, and evaluates its effect on outcome and quality.

As mentioned in Section 3.1, the perishable aspect of inventory is modelled by the transition of goods from a fresh state (Fresh Finished Goods Inventory) to an aged state (Previous Finished Goods Inventory). The model differentiates between inventory in only two ways, fresh and aged, based on their impact on the system. The system dynamics methodology is an efficient means of monitoring and analyzing such variables for this purpose. In mathematical models, inventory is represented in a very exact manner, indexed by time of production and sale [28], [31]. This can take the focus away from the more important implications of inventory.

Four scenarios are simulated: FIFO applied to Perishables, FIFO applied to Non-Perishables, LIFO applied to Perishables, and LIFO applied to Non-Perishables. Each scenario is evaluated by its resulting inventory level. The analysis can be segmented into three parts: (1) A comparison between the Perishables and Non-Perishables scenarios for logic-checking, (2) A comparison of the FIFO and LIFO scenarios for policy evaluation, and (3) Determining the optimal issuance policy to apply to perishable inventories.

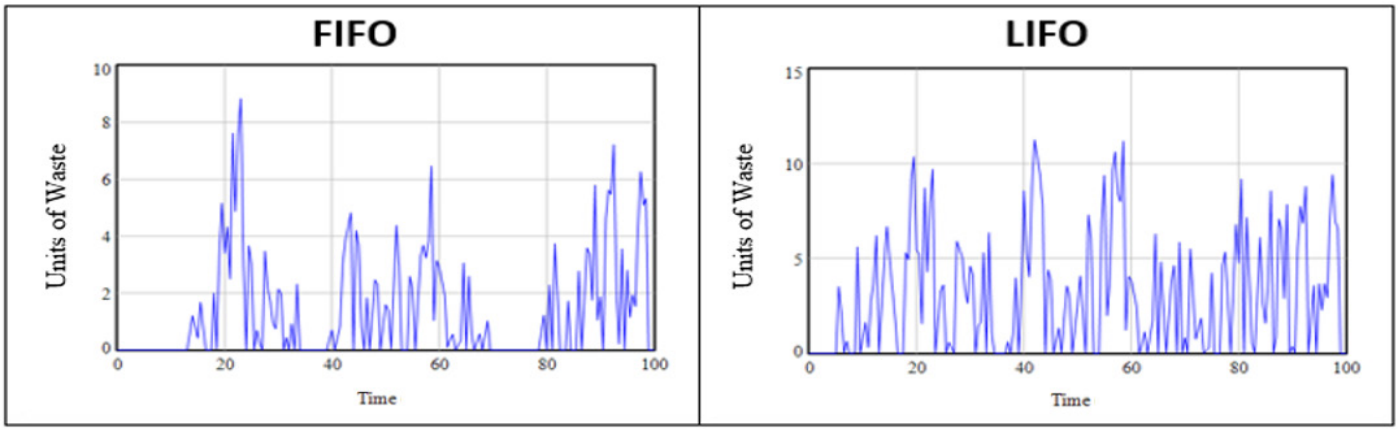
4.1. Comparison of Perishables and Non-Perishables Scenarios

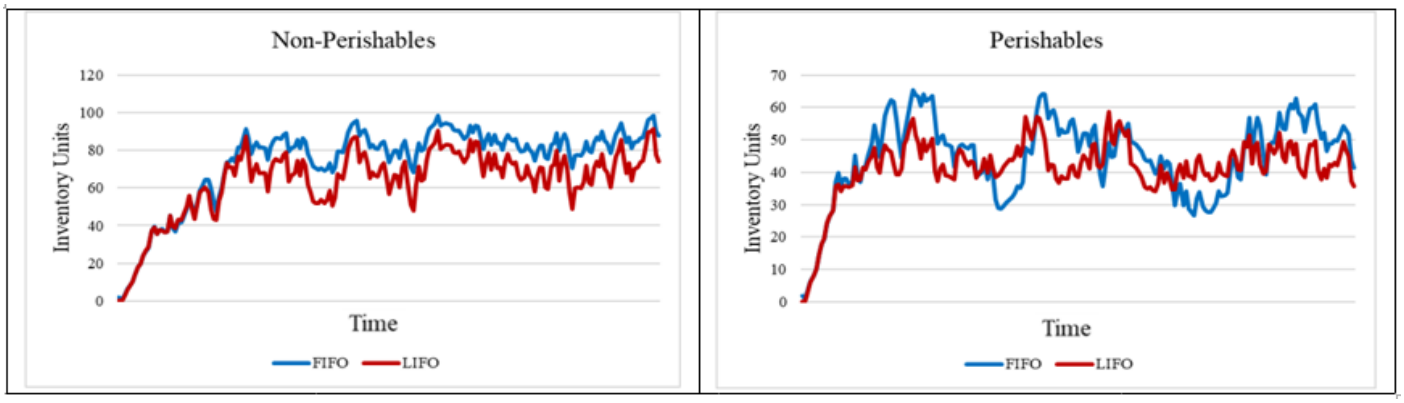
In Figure 5, sections a and c contain the simulated inventory levels for the perishables scenario, while sections b and d contain the results for non-perishables. In all cases, inventory levels exhibit oscillating behavior which is aligned with the behavior being demonstrated in other inventory models and supply chain industry studies [36]. The perishables and non-perishables scenarios differ in the long-term trend being demonstrated by inventory under each case. Under the non-perishables scenario, inventories exhibit an upwards trend, indicating the tendency for inventories to accumulate. Under the perishables scenario, inventory levels are stable in the sense that they exhibit no upwards or downwards trend. The volume of inventory in the perishables scenario is naturally regulated by the potential for wastes, which prevents accumulation. On the other hand, there is no negative feedback to control the volume of inventories in the non-perishables scenario, resulting in accumulation.

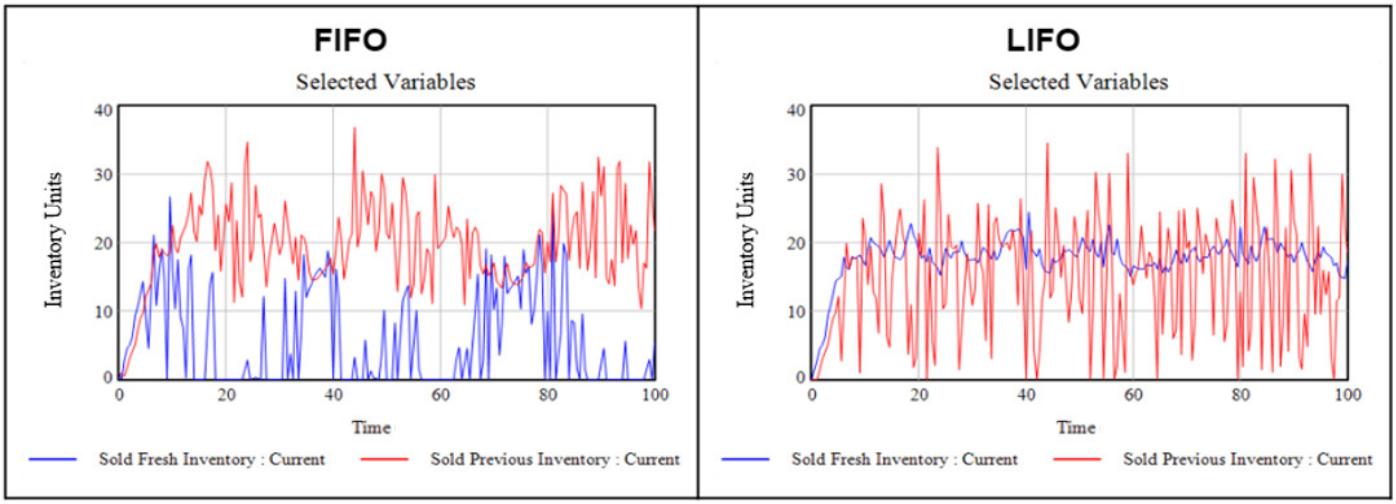
The simulated behavior for non-perishables is aligned with the behavior of most forecast-based non-perishable inventory systems. There is evidence to support this in the management of mechanical parts inventory, which has a large amount of excess [38], [39]. The accumulation may also be attributed to the fact that production quantity was based solely on sales forecasts, without any regard for current inventories. This mimics decision-making for perishables more closely. Since goods manufactured in one period do not have the same external value as goods manufactured in a different period, past inventory cannot be fully utilized to supply current demand. Therefore, decision-making parameters such as current inventory and safety stock generally carry more weight in models for non-perishable inventories.

The simulated behavior for perishable inventories follows the behavior of traditional inventory models (e.g. EOQ), which oscillate without any positive or negative trend. Inventory accumulates because of inaccuracies in forecast. Yet, because a fraction of perishable inventories become waste at the end of each period, inventories cannot accumulate. Instead, excess inventory from inaccuracies in in forecasting eventually leave the system as waste. This is not favorable given that the amount of inventory that become Wastes are just excess in a different form. It may only be said that inventory is more effectively-handled under the perishable scenario if the volume of its Wastes is not increasing in the same way excess inventory does for non-perishables. The simulated behaviors for waste, although oscillating, exhibit no increasing trend (see Figure 6). For these reasons, it may be concluded that perishable inventories do not have as strong of a tendency towards overproduction.

It may be concluded based on the simulated results for Previous Finished Goods Inventory that the main issue in the management of perishable inventory is not preventing surpluses and shortages. Under common inventory management policies, perishability as a product characteristic results in an inventory system that is less







prone to such outcomes. Inventory models with the objective of determining an optimal stockout level could thus be redundant, or even counterproductive to the system's natural ability to regulate itself. On the other hand, the system could benefit from the minimization of its peak inventory levels, or all-around variability.

As the simulated behaviors follow the trend described by real industry data and by literature, the ensuing analysis can be deemed valid. A comparison may now be conducted between the FIFO and LIFO policies given the system structure. Based on the simulated values in Figure 5, it is apparent that the choice of issuance policy hardly makes a difference on inventory levels. On the other hand, there are clear differences between the FIFO and LIFO scenarios in the perishables case.

4.2. Comparison of FIFO and LIFO Scenarios

Generally, inventory levels are higher under the FIFO case. This is true for both the non-perishables and perishables cases. Figure 7 shows that in, the non-perishables case, inventory levels

exhibit the same upwards trend. Yet, once the values stabilize, the FIFO case will begin to exhibit consistently higher inventory levels. Its highest point and lowest point upon stabilization are higher than inventory at any point in time under LIFO. This is different from the case of perishables, where inventory levels have higher maximum peaks under FIFO but also lower LIFO peaks. However, based on the simulations, it would appear that FIFO exhibits higher inventory levels more often than it exhibits lower inventory levels.

The higher inventories under FIFO may be attributed to the extent of feedback delay between production and demand. In the causal loop diagrams in Figure 2, there are multiple delays between Fresh Finished Goods Inventory and Demand under FIFO. This signifies that there are multiple delays between demand and production for forecasts based on it are actually sold. Hence, there is a higher propensity for overproduction or underproduction. In these cases, the problem is leaning towards overproduction given that current inventories are not considered in deciding on the quantity to produce in the next period.

Based on the review of studies on PIS, there are more models were FIFO proved to be optimal than otherwise. These models are optimization models with the primary objective of minimizing costs. The potential for waste production is incorporated in the form of disposal costs or deterioration costs, which align the sustainability and economic objectives for the system. It is understandable for FIFO to be optimal following this perspective, as FIFO results in the lower minimum inventory values. However, in actual inventory systems, stable inventories are preferred over fluctuations.

In literature, the case where LIFO proved to be optimal was the case where prices decreased linearly for each period that the inventory unit remained in the system. This would suggest that when quality affects revenues, LIFO becomes preferable. This is because the profit earned per unit would be more consistent. The key benefit to be derived from LIFO is consistency. In this model, quality affects revenues by acting as a multiplier to demand. Results showed that inventory levels were less variable for both perishables and non-perishables, which aligns with the suppositions from previous studies that LIFO can result in greater consistency.

4.3. Evaluation of FIFO and LIFO Applied to Perishables

For perishables, product quality is partially determined by the age of inventories. Hence, the volume of fresh and aged inventories can have an impact on demand. In this model, quality is given as the fraction of fresh inventories in relation to total inventories. Previously, it was mentioned that the levels of fresh and aged inventories that result from FIFO and LIFO respectively different greatly for perishables. This means that the use of FIFO and LIFO can also significantly alter the quality of perishable goods being sold.

It has been mentioned that LIFO may promote greater consistency in an inventory system. This also manifests in the current model where quality is more consistent under LIFO due to the fact that the levels of both fresh and aged inventory are more consistent. Consistent quality also results in greater consistency of sales. Figure 8 contains the simulated values for sales of inventory, where it may be observed that Sold Fresh Inventory, which is the primary contributor to sales in LIFO, is high consistent.

Under FIFO, sales of fresh and aged inventories both fluctuate. Specifically, the two inventories have an inverse relationship, as the sales that cannot be met with aged inventories will satisfied with fresh inventories. The fluctuations in quality also lead to fluctuations in demand, which feedbacks to inventory levels. When demand is high, it can prompt the production of new inventory, which has a negative effect on demand as the goods will be aged by the time they are sold. This problem is aggravated by the fact that there the production quantities derived from past demand will be misaligned with actual demand. When demand is rising, production will be higher, and conversely, when demand is dropping, production will be lower. This is the result of feedback delays between demand and production.

There are also feedback delays upstream, in the customer's perception of quality. The current model differentiates between actual quality and perceived quality, by the logic that there is a delay in the recognition of quality change when it comes to inventories. In most inventory models, the perception of quality and its resulting change in demand is instantaneous. However, this relationship is only descriptive of settings such as retail stores where customers see the physical units individually and its quality instantaneously affects the probability of purchase [8], [40]. This is not the case for entities where purchases are made in bulk, such as the manufacturing or warehousing settings. In such cases, the individual quality of units have little value in the system, which is evident by the miss-outs from sampling procedures [3]. There are also delays due to the fact that not all inventory units are assessed prior to production, and thus any information on quality would only be known after production or through customer returns [41].

FIFO exhibits longer cycles between the arrival and sale of an inventory unit. In the short term, this will result in (1) less wastes and lower inventory levels for the manufacturing firm, and (2) instances when sales may be lower in the periods during and following the end of an inventory cycle, when the leftover units have aged. The trade-offs can be managed with a classical optimization model that can balance waste generation with freshness. On the other hand, the long-term impact of FIFO in the supply chain is inconsistent quality. This complicates reordering decisions downstream in the supply chain, by adding variability to product lifetimes. This in turn heightens the amplitudes of bullwhips being exhibited by the inventory levels of a supply chain. Therefore, in the long term, FIFO promotes variability and greater efforts being made by members of a supply chain to manage it.

LIFO has the opposite effect on inventory levels in the short term; and will result in more consistent quality in the long term. While this lessens the complexity of decision-making for the members of the supply chain, there is nonetheless the issue of waste to be addressed. The consistent generation of waste would require greater efforts in waste management. On a positive note, the consistent waste generation would also regularize the demand for waste management, and through that require the standardization of waste management procedures. Currently, the efforts to minimize waste and the natural variability of production add complexity to decisions on when and where to set up waste management

facilities. Given the high investment required by such facilities, this generally results in there being insufficient infrastructure for this purpose. In this sense, the consistency promoted by implementations of LIFO would be beneficial, by giving a clearer perspective on the volume of waste management activity that would be needed.

5. Conclusion

In this study, classic FIFO and LIFO issuance policies were assessed for their impact in a perishable inventory system. The performance of these policies was evaluated in terms of their ability to meet sustainability and quality goals. Specifically, it aimed to improve sustainability through waste reduction and quality by maximizing the freshness of inventories in the system. While similar studies had been made in the past, the present study viewed the problem from a different approach, through the development of a system dynamics (SD) framework. Using the SD framework emphasizes the patterns of the inventory levels, rather than its exact values, allowing for the generation of insights that can be generalized regardless of lifetime variability.

A simulation model was developed comprising of three major sections that eventually form a general single loop: (1) The physical inventory system, (2) The perception of quality, and (3) The perception of demand. Furthermore, feedback delays such as delays in perishing, perceived quality, and perceived demand were also considered. FIFO and LIFO models for both perishables and non-perishables products were then formed with the presence of Waste outflow variable as a distinguishing factor between non-perishables and perishables model.

Comparing the perishables and non-perishables scenario yielded that for all cases, inventory levels exhibit an oscillating behavior. Non-perishables showed upward trend due to accumulation of inventory since quantity to be produced relies solely on sales forecasts and that current leftover usable inventory was not factored in as to how much quantity is to be produced for the next period. This is to emulate perishables inventory systems wherein external values of goods differ per period. On the other hand, perishables seem to have no upward for downward trend and seem to follow the behavior of traditional inventory models like EOQ. Perishable inventories do not have strong tendency towards overproduction. Wastes is not increasing the same way excess inventory does for non perishables. Based on these observations, it may be counterproductive to determine the optimal stockout level of perishable goods since it seems to have its way of regulating itself.

A comparison of the FIFO and LIFO scenarios showed that LIFO is preferable when revenues are heavily influenced by quality since LIFO promotes greater consistency in inventory system. Consistency in quality brings consistency in profit earned per unit and this was explicitly observed in Sold Fresh Inventory which is a primary contributor to the sales in LIFO.

On the other hand, FIFO exhibits higher inventory levels which is attributed due to feedback delays between production and demand. Furthermore, FIFO system tends to lean towards overproduction since current leftover inventories were not factored into the decision for the amount or quantity to be produced for the next period. Also, fluctuations were observed in FIFO for both sales in fresh and aged inventories - as evidenced by its inverse relationship - lead to fluctuations in demand. Although FIFO may be more optimal for minimizing costs and inventory levels once wastes is treated as disposal costs or deterioration costs, actual inventory systems still prefer stable inventories rather than constant fluctuations. FIFO is preferable for cost-sensitive or waste-sensitive cases, while LIFO is preferable in quality-sensitive cases.

The results of this study demonstrate the ability of SD to produce insights in addition to those that can be derived from optimization models. While the current study applied SD on classic inventory management policies (i.e. FIFO and LIFO), literature on perishable inventory models comprise various new policies and technologies that can be assessed more extensively using the SD framework. In particular, information technologies such as RFID and IoT have been proposed as means of monitoring the age-based quality of perishables with better accuracy, yet there is still the question of how to use this feedback effectively in decision-making. The current study serves as a starting point for future studies applying SD in evaluating and improving policies that will prove useful in a wide variety of cases.

One of the major contributions of this study is its demonstration of the SD framework as a means of modelling and simulating policy outcomes that are difficult to, or cannot be specifically quantified. This is very relevant to the field as it recent literature centers on the application of information technology in inventory management. The effects of greater accuracy and shortened feedback granted by information technology are not directly quantifiable, making it a suitable extension of this policy. The study also contributes by providing insight on what the long-term implications of issuance policies could be, particularly in the area of sustainability. While it may appear that FIFO is the most efficient in terms of cost and waste generation, LIFO actually has more positive long-term implications by reducing complexity in supply chain decisions, and standardizing waste management efforts.

References

[1] J. Blackburn and G. Scudder, “Supply Chain Strategies for Perishable Products : The Case of Fresh Produce,” Prod. Oper. Manag., vol. 18, no. 2, pp. 129–137, 2009.

[2] G. Ferreira, E. Arruda, and L. Marujo, “Inventory management of perishable items in long-term humanitarian operations using Markov Decision Processes,” Int. J. Disaster Risk Reduct., vol. 31, pp. 460–469, 2018.

[3] Y. Yang, H. Chi, O. Tang, W. Zhou, and T. Fan, “Cross perishable effect on optimal inventory preservation control,” Eur. J. Oper. Res., vol. 276, no. 3, pp. 998–1012, 2019.

[4] M. Dehghani and B. Abbasi, “An age-based lateral-transshipment policy for perishable items,” Int. J. Prod. Econ., vol. 198, pp. 93–103, 2018.

[5] L. Janssen, T. Claus, and J. Sauer, “Literature review of deteriorating inventory models by key topics from 2012 to 2015,” Intern. J. Prod. Econ., vol. 182, pp. 86–112, 2016.

[6] C. Kouki, M. Z. Babai, and S. Minner, “International Journal of Production Economics On the bene fi t of dual-sourcing in managing perishable inventory,” Intern. J. Prod. Econ., vol. 204, no. May, pp. 1–17, 2018.

[7] Z. Azadi, S. D. Eksioglu, B. Eksioglu, and G. Palak, “Stochastic optimization models for joint pricing and inventory replenishment of perishable products,” Comput. Ind. Eng., no. November, pp. 1–18, 2018.

[8] B. Afshar-nadja, “The influence of sale announcement on the optimal policy of an inventory system with perishable items,” J. Retail. Consum. Serv., vol. 31, no. C, pp. 239–245, 2016.

[9] X. Xu, Q. Bai, and M. Chen, “two-warehouse inventory systems for deteriorating items,” Appl. Math. Model., vol. 0, pp. 1–16, 2016.

[10] L. Janssen, Development and simulation analysis of a new perishable inventory model with a closing days constraint under non-stationary stochastic demand. 2018.
[11] L. Janssen, A. Diabat, J. Sauer, and F. Herrmann, "A stochastic micro-periodic age-based inventory replenishment policy for perishable goods," Transp. Res. Part E, vol. 118, no. January, pp. 445–465, 2018.
[12] Y. Lee, S. Mu, Z. Shen, and M. Dessouky, "Computers & Industrial Engineering Issuing for perishable inventory management with a minimum inventory volume constraint," Comput. Ind. Eng., vol. 76, pp. 280–291, 2014.
[13] G. Gaukler, M. Ketzenberg, and V. Salin, "International Journal of Production Economics Establishing dynamic expiration dates for perishables : An application of RFID and sensor technology," Int. J. Prod. Econ., vol. 193, no. November 2016, pp. 617–632, 2017.
[14] A. Hiassat, A. Diabat, and I. Rahwan, "A genetic algorithm approach for location-inventory-routing problem with perishable products," J. Manuf. Syst., vol. 42, pp. 93–103, 2017.
[15] R. Dominguez, B. Ponte, S. Cannella, and J. M. Framinan, "On the dynamics of closed-loop supply chains with capacity constraints," Comput. Ind. Eng., 2018.
[16] M. Song, X. Cui, and S. Wang, "International Journal of Production Economics Simulation of land green supply chain based on system dynamics and policy optimization," Intern. J. Prod. Econ., no. July, pp. 0–1, 2018.
[17] X. Wang, S. M. Disney, and J. Wang, "Int . J . Production Economics Exploring the oscillatory dynamics of a forbidden returns inventory system," Intern. J. Prod. Econ., vol. 147, pp. 3–12, 2014.
[18] M. R. S. M, C. Hora, D. O. Fontes, F. Gaudêncio, and M. Freires, "Sustainable and renewable energy supply chain : A system dynamics overview," Renew. Sustain. Energy Rev., vol. 82, no. June 2017, pp. 247–259, 2018.
[19] T. Rebs, M. Brandenburg, and S. Seuring, "System dynamics modeling for sustainable supply chain management : A literature review and systems thinking approach," J. Clean. Prod., 2018.
[20] M. Kalverkamp and S. B. Young, "In Support of Open-Loop Supply Chains :," J. Clean. Prod., 2019.
[21] H. Jalali, R. Carmen, I. Van Nieuwenhuyse, and R. Boute, "Quality and Pricing Decisions in Production/Inventory Systems," Eur. J. Oper. Res., vol. 272, no. 1, pp. 195–206, 2018.
[22] S. Kumar and A. Nigmatullin, "Simulation Modelling Practice and Theory A system dynamics analysis of food supply chains – Case study with non-perishable products," Simul. Model. Pract. Theory, vol. 19, no. 10, pp. 2151–2168, 2013.
[23] Z. Hosseinifard and B. Abbasi, "Computers & Operations Research The inventory centralization impacts on sustainability of the blood supply chain," Comput. Oper. Res., pp. 1–7, 2016.
[24] P. M. Ching and J. A. Wu, "An Assessment of FIFO and LIFO Policies for Perishable Inventory Systems Using the System Dynamics Approach," in 2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA), 2019, pp. 251–255.
[25] R. Poles, "Int . J . Production Economics System Dynamics modelling of a production and inventory system for remanufacturing to evaluate system improvement strategies," Intern. J. Prod. Econ., vol. 144, no. 1, pp. 189–199, 2013.
[26] Y. Qiu, J. Qiao, and P. M. Pardalos, "Optimal production, replenishment, delivery, routing and inventory management policies for products with perishable inventory," Omega, vol. 82, no. C, pp. 193–204, 2019.
[27] P. Moser, O. Isaksson, and R. W. Seifert, "Inventory Dynamics in Process Industries : AC SC," Int. J. Prod. Econ., 2017.
[28] J.-S. Song and Y. Zhao, "The Value of Component Commonality in a Dynamic Inventory System with Lead Times," Manuf. Serv. Oper. Manag., vol. 11, no. 3, pp. 493–508, 2009.
[29] M. Gaggero and F. Tonelli, "ScienceDirect Optimal Control of Distribution Optimal Control of Distribution Optimal of Distribution Perishable Goods Optimal Control Control of Distribution Perishable Goods Perishable Goods Perishable Goods Chains Chains Chains Chains for," IFAC-PapersOnLine, vol. 48, no. 3, pp. 1049–1054, 2015.
[30] Y. Duan, G. Li, J. M. Tien, and J. Huo, "Inventory models for perishable items with inventory level dependent demand rate," Appl. Math. Model., vol. 36, no. 10, pp. 5015–5028, 2012.
[31] A. A. Alamri and A. A. Syntetos, "Beyond LIFO and FIFO : Exploring an Allocation-In-Fraction-Out (AIFO) policy in a two-warehouse inventory model," Intern. J. Prod. Econ., vol. 206, no. September, pp. 33–45, 2018.
[32] C. C. Lee, "Two-warehouse inventory model with deterioration under FIFO dispatching policy," vol. 174, pp. 861–873, 2006.
[33] Y. Li, F. Chu, Z. Yang, and R. W. Calvo, "A Production Inventory Routing Planning for Perishable Food with Quality Consideration," IFAC-PapersOnLine, vol. 49, no. 3, pp. 407–412, 2016.
[34] J. Huber, A. Gossmann, and H. Stuckenschmidt, "Cluster-based hierarchical demand forecasting for perishable goods," Expert Syst. Appl., vol. 76, pp. 140–151, 2017.
[35] L. A. San-josé, J. Sicilia, and J. García-laguna, "Optimal lot size for a production – inventory system with partial backlogging and mixture of dispatching policies," Intern. J. Prod. Econ., pp. 1–10, 2013.
[36] A. C. Braz, A. M. De Mello, L. Augusto, D. V. Gomes, P. Tromboni, and D. S. Nascimento, "The bullwhip effect in closed-loop supply chains : A systematic literature review," J. Clean. Prod., vol. 202, pp. 376–389, 2018.
[37] G. Dobson, E. J. Pinker, and O. Yildiz, "An EOQ model for perishable goods with age-dependent demand rate," Eur. J. Oper. Res., vol. 257, no. 1, pp. 84–88, 2017.
[38] Q. Hu, J. E. Boylan, H. Chen, and A. Labib, "OR in spare parts management : A review," Eur. J. Oper. Res., vol. 266, no. 2, pp. 395–414, 2018.
[39] S. Rahimi-ghahroodi, A. Al Hanbali, I. M. H. Vliegen, and M. A. Cohen, "International Journal of Production Economics Joint optimization of spare parts inventory and service engineers staffing with full backlogging," Intern. J. Prod. Econ., vol. 212, no. February, pp. 39–50, 2019.
[40] H. Hvolby and K. Steger-jensen, "Managing cannibalization of perishable food products in the retail sector," Procedia - Procedia Comput. Sci., vol. 64, pp. 1051–1056, 2015.
[41] K. Hun and H. Hwang, "A returns policy for distribution channel coordination of perishable items," vol. 152, pp. 770–780, 2004.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 82-87 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

A Study on Development of Evaluation Metrics for Learners in Physical Computing

Woochun Jun*

Seoul National University of Education, Department of Computer Education, 06639, Korea

ARTICLE INFO

Article history:
Received: 16 June, 2019
Accepted: 11 August, 2019
Online: 03 September, 2019

Keywords:
Physical Computing
Evaluation Metrics
ICT Education
Software Education

ABSTRACT

Physical computing is important for ICT (information and communication technology) Education and other informatics education such as software education since physical computing can provide learning-by-doing education for students. It is also a strong tool to increase students' programming ability using various type of physical computing tools like a robot. In physical computing, it is necessary to have evaluation metrics for learners in order to test students' progresses and environmental safety, etc. However, in the literature, there is little work for evaluation metrics for learners. In this paper, for the proposed evaluation metrics for learners developed in the previous research, the usefulness and validity are discussed and justified with statistical analysis. The final 23 evaluation metric will be very useful in actual physical computing education.

1. Introduction

This paper is an extension of work originally presented in ICTC2018[1].

Physical computing refers to downloading data from the real world to a digital device, processing it in the form of software, and then outputting the results to a monitor, LED or various devices. To put it more simply, computers and the real world are talking through data. The main content of the physical computing was developed by artists dealing with New Media Art. This is because it was necessary to understand computers in order to create works in the field of media art using images or light.

Physical computing is the foundation of the Internet of Things field. The basic nature of the Internet of Things is that it recognizes its surroundings through sensors and shares or processes the necessary information over the network, so it can be said to be an important technology that forms the Internet of Things.

On the other hand, for students who are new to programming, physical computing can be a good way to learn how computers and programs work. For example, students can learn about computers by assembling blocks directly, controlling robots by assembling and controlling them, or using sensors and outputs that can be easily detached to measure the quality and temperature of air which makes it easier for them to understand the principles of machine and programming.

Simplified coding programs and computer devices help students learn about physical computing. Instead of advanced programming languages such as C, students can learn programming simply by using block-coding software called Scratch. And instead of dealing with complex computer parts, students can understand the principles of computers by assembling microcomputer boards such as Arduino and Raspberry Pies, as well as various sensors and outputs to complete their own machine.

With this characteristic that the virtual world and the real world can exchange information, it is expected that the software industry will play an important role in the future thanks to the fact that it is the basis of the Internet of Things field that recognizes the surrounding environment and processes the necessary information, and that it can be approached with a simple and active image in terms of programming education.

In physical computing, evaluation metrics for learners are very important. Those metrics are very useful in evaluating learners' progresses, teaching-learning models, class materials, etc. However, in the literature, there is little research work on evaluation of learners in physical computing. The purpose of this paper is to develop evaluation metrics of learners in physical computing and justify them based on rigorous statistical analysis.

The rest of this paper is organized as follows. In Chapter 2, related works for physical computing are introduced. In Chapter3, evaluation metrics of learners are proposed and discussed. In Chapter 4, statistical analysis is presented for those evaluation

*Corresponding Author: Woochun Jun, Seocho Dong, Seocho Gu, Seoul, Korea, wocjun@snue.ac.kr

metrics. Finally, in Chapter 5, conclusions and further research issues are presented.

2. Related Works

2.1. Evaluation standards of Physical Computing

In this section, literature review works for evaluation standards for robot computing and physical computing are introduce.

In [2], achievement criteria of robot computing are proposed as below. In their work, criteria of robot computing is classified into 5 categories, understanding of robot, robot experience, robot motor control and sensor experience, control of sensor and creation of works, project in daily life, respectively.

[Step 1] Understanding Robot

(1) Definition of robots

I can understand and explain the basic elements of robot definition.

(2) Robot type and composition

I can understand the kinds of robots used in real life and the parts and applications that make up them.

(3) Robot operating principles

Robots understand how they react through external input or recognition and know how they work.

[Step 2] Robot Experience

(4) Robotic rules and safe use

I can understand the basic rules for building robots and how to use them safely.

(5) Understanding robot behavior

It is able to know necessary procedures for robot operation such as connecting devices and turning them on with variety of robots.

(6) Fabricate simple motion robot

Robotic behavior can be created in the order presented through simple block programming. It can also be created by modifying robot movements by applying various themes.

(7) Brief description of operation robot

Simple block programming can create robot behavior and explain the contents of the procedure.

[Step 3] Robot Motor Control and Sensor Experience

(8) Understanding Rotational Motion

Understanding the motor, learning about the operating procedure, principle of rotation, and component control required for member operation can be understood and simple robot rotation operation can be created and understood.

(9) Various robot-driven work production

Based on the principle of motion and understanding of the robot's driving procedure, various robots can be used to make a drive. I can also learn how to control the precise movement of a robot.

(10) Utilization of robot drive works

I can explain to my friends how the robot drives, how the robot is programmed in order, and its functions. It develops social relationship cooperation and communication skills through sharing of works and expression activities (recognition of goal achievement, sharing of error-solving experiences, etc.).

(11) Creation of simple sensor robot works

I can create a robot or device that can control behavior combined with simple sensors. It can make robots or devices using one of several sensors, including light sensors, ultrasonic sensors and sound sensors.

(12) Brief description of sensor robot work

I can identify the types of sensors used in my life, understand the characteristics of sensors, and describe robots or devices with sensors. I can express to my friends the design, function and programming procedures of a robot or device-driven work combined with one sensor. It develops social relationship cooperation and communication skills through sharing of works and expression activities (recognition of goal achievement, sharing of error-solving experiences, etc.).

[Step 4] Control of Sensors and Creation of Works

(13) Produce various sensor robot works

I can create and program controllable robot or device works by combining various sensors that can solve problems found in my life. It can also create works and explain them logically according to the programming sequence according to the purpose and purpose of production.

(14) Description of the various sensor robot works

I can describe the function, use and programming procedures and methods of robot or device work combined with various sensors. It also develops social relationship cooperation and communication skills through works sharing and expression activities (recognition of goal achievement, sharing of error-solving experiences, etc.).

(15) Rule Design and Robot Fabrication

Understanding the conditions for compliance with the rules, understanding the algorithms that can perform them, and designing the parts and devices needed to drive can be designed to operate through programming that applies algorithms to create robots or devices and perform them.

[Step 5] Project in Daily Life

(16) Designing Robot Works in Life

I can think about topics that can solve life problems and design algorithms and creative works that can define necessary parts and functions and control them by individuals or teams.

(17) Making Robot Works in Life

A creative work can be created through programming with algorithms applied to create robots or devices designed by individuals or teams and to perform defined functions according to topics that can solve life problems.

(18) Sharing and Expression of Robot Works

To solve problems found in life, creative works created by individuals or teams can be shared with friends, and better improvements can be found and expressed. In addition, better algorithms can be found, supplemented and shared to perform functions.

2.2. Educational Effects of Physical Computing

In [3], they presented EPL and physical computing curriculum using Entry and sensor boards that can be easily learned and used by elementary school students. The results of this study showed that both the comparative groups applying the problem-oriented EPL curriculum and the groups applying the additional problem-oriented EPL curriculum and the physical computing curriculum have significant effects on the improvement of students' computational thinking skills. Specifically, the problem-oriented EPL curriculum and the physical computing curriculum using the sensor board contributed to increasing learner's problem-solving skills, and showed that they are more effective in terms of improving satisfaction and computational thinking-related problem-solving skills when the physical computing education is combined with the programming training alone

In [4], high school students in the industrial field were trained in programming Android applications that control Arduino using app computers and analyzed the changes in students' creative and fusion thinking abilities. To this end, they designed a teaching course based on a creative problem-solving model and an integrated thinking model. Studies have resulted in significant improvements in diffuse thinking and motivational factors among the subcomponents of creative problem solving. In addition, according to the results of a survey by students on integrated thinking, many students answered that they can devise IoT systems that can be applied to their daily lives based on the knowledge learned through this class. Thus, the education of physical computing using app computers and Aduinos could be seen to have a positive effect on students' ability to think creatively and diversely.

On the other hand, other research works showed benefits of physical computing in ICT education [5-7].

3. Development of Evaluation Metrics for Learners

3.1 Design Principles

In order to development of evaluation standards of learners in physical computing, the following principles are adopted [1].

First, student-oriented activity is emphasized. Second, process-oriented activity is emphasized. Third, cooperative activity is encouraged rather than individual study in physical computing. Fourth, attitudinal and emotional experience or achievement are included.

In overall, evaluation standards are developed based on educational philosophy called constructivism [8]. Constructivism has the following major characteristics. First, it is assumed that there is no objective knowledge in the real-world. It means that every can make his or her own knowledge depending on his or her way of thinking or view. Second, it encourages social interaction. In other words, cooperative works are encouraged for study. Third, it also emphasizes problem-solving ability for authentic problems. Authentic problems mean that problems can happen in our daily life.

3.2 Evaluation Metrics

The following evaluation standards of physical computing are proposed in [1].

Table 1. Evaluation Standards of Learners

Division	Standards
Study Guidance	-Do students know study objective in advance or in the beginning of the activity? -Do students know any prerequisite or background information necessary for the activity? -Do students know the necessary equipment or teaching aids for the activity? -Do students know study materials for the physical computing activity?
Cooperation	-Does each student have a proper role in physical computing activity? -Is any leader in the group? -Are ways of communication good enough during activity?
Safety	-Do students know safety instruction for the activity? -Do students break or ruin any parts in the activity?
Output	-Do students finish the activity and produce output? -Is output good enough as expected or incomplete? -Are students supposed to have a change to expression their review or impression of the activity?
Report	-Do students write a memo or report during the activity? -Do students produce a final report after end of the activity?
Attitude	-Are students polite in the process of activity? -Does any student cause a brawl and stop the activity?
Time	-Do students spend right time for the whole class? -Do students spend too much time in specific activity of the class?
Emotion	-Are students satisfied with the activity? -Are students interested in physical computing activity? -Do students have some confidence on the subject after activity?
Theory	-Do students know the principle or theory of the physical computing activity? -Can students link the theory with physical computing activity?

4 Statistical Analysis

4.1 Design of Analysis

Verification of the reliability of the tool to measure the importance of evaluation metrics for learners in physical computing showed that Cronbach α was 0.90 and was reliable, as shown in Table 2.

Table 2. Reliability of Measurement Tool

Item Number	Cronbach's α
23	0.90

The data collected in this study were analyzed using the Statistical Package for the Social Science (SPSS) WIN 25.0 program. The analyzer method calculated Cronbach α to verify the reliability of the measuring instrument. In addition, the average and standard deviation were obtained in order to find out the importance of the evaluation standards of learners in the physical computing, and the correlation was conducted to understand the relationship between the sub-areas of the evaluation metrics of learners in the physical computing.

4.2 Analysis Results

-Descriptive Statistics

The average and standard deviation of questions asked about the importance of evaluation metrics of learners in physical computing are as shown in Table 3.

Table 3. Results of Descriptive Statistics

Evaluation Standard	AVG.	SD
1. Do students know study objective in advance or in the beginning of the activity?	3.66	1.10
2. Do students know any prerequisite or background information necessary for the activity?	3.16	1.14
3. Do students know the necessary equipment or teaching aids for the activity?	2.91	1.17
4. Do students know study materials for the physical computing activity?	2.84	1.11
5. Does each student have a proper role in physical computing activity?	3,88	1.04
6. Is any leader in the group?	3.22	1.21
7. Are ways of communication good enough during activity?	4.22	1.01
8. Do students know safety instruction for the activity?	4.56	0.72
9. Do students break or ruin any parts in the activity?	3.50	1.30
10. Do students finish the activity and produce output?	3,41	0.95
11. Is output good enough as expected or incomplete?	3.16	0.92
12. Are students supposed to have a change to expression their review or impression of the activity?	4.06	0.91
13. Do students write a memo or report during the activity?	2.97	1.23
14. Do students produce a final report after end of the activity?	2.94	1.19
15. Are student polite in the process of activity?	4.13	0.91
16. Does any student cause a brawl and stop the activity?	3.69	0.90
17. Do students spend right time for the whole class?	4.34	0.70
18. Do students spend too much time in specific activity of the class?	3.31	1.09
19. Are students satisfied with the activity?	4.28	0.68
20. Are students interested in physical computing activity?	4.41	0.67
21. Do students have some confidence on the subject after activity?	4.19	0.93
22. Do students know the principle or theory of the physical computing activity?	3.56	1.11
23. Can students link the theory with physical computing activity?	3.66	1.04

Where AVG and SD stand for average and standard deviation, respectively.

4.3 Sub-area of Descriptive Statistics

The results of the average and standard deviation of the sub-area of the importance of the evaluation metrics of learners in physical computing showed that the average of the importance of the evaluation metrics of learners in physical computing was 3.65, as shown in Table 4.

Among the sub-categories of the importance of learners' evaluation metrics for physical computing, the average was 4.29, followed by "safety" 4.03, "attitude" 3.91, "time" 3.83, "cooperation" 3.77, "theory" 3.61, "output" 3.54 and "student guidance" 3.14 and "report" 2.95 were the lowest.

Table 4. Results of Descriptive Statistics of Sub-area

Evaluation Standard Area	AVG.	SD
Student Guidance	3.14	0.97
Cooperation	3.77	0.92
Safety	4.03	0.86
Output	3.54	0.67
Report	2.95	1.17
Attitude	3.91	0.78
Time	3.83	0.68
Emotion	4.29	0.58
Theory	3.61	1.02
Overall Evaluation Metrics	3.65	0.56

Where AVG and SD stand for average and standard deviation, respectively.

As shown above in Table 4, among the sub-areas of the learners' evaluation metrics in physical computing, the importance of 'emotion' is highest and the importance of 'report' is lower than that of other areas.

4.4 Correlation Analysis

The results of the correlation between the sub-area of evaluation metrics of learners in physical computing and the entire evaluation metrics are shown in Table 5.

Where * p<.05, *** p<.001

As shown above in Table 5, student guide(r=.661, p<.001), cooperation(r=.802, p<.001), safety(r=.716, p<.001), output(r=.419, p<.05), report (r=.764, p<.001), attitude(r=.696, p<.001, time(r=.614, p<.001), emotion(r=.616, p<.001), and theory (r=.606, p<.001) showed a statistically significant static correlation with evaluation metrics of learners in physical computing. Therefore, the higher the importance of student guidance, cooperation, safety, output, report, attitude, time, emotion, and theory, the higher the importance of evaluation metrics of learners in physical computing.

Table 5. Results of Correlation Analysis

Area	Correlation Value
Student Guidance	0.661***(0.000)
Cooperation	0.802***(0.000)
Safety	0.716***(0.000)
Output	0.419*(0.017)
Report	0.764***(0.000)
Attitude	0.696***(0.000)
Time	0.614***(0.000)
Emotion	0.616***(0.000)
Theory	0.606***(0.000)

5 Conclusions and Further Research Issues

Physical computing is very important in current ICT education and software education since it can provide learning-by-doing activity for learners. In other words, learners can enjoy the physical computing activity and have interests and motivation in physical computing class. Physical computing becomes very popular as ICT education is extended in the future.

In physical computing, evaluating learners is very essential in checking learners' study progress. Also, evaluating learners is also important in developing teaching-learning models in classrooms, study contents, etc. In the literature, there is little work on evaluation standards or metrics of learners. In the previous research work, new evaluation standards are developed and proposed. Through statistical analysis, overall evaluation standards are justified to use in physical computing.

In this research, based on previous research work, the purpose of the work is to justify the results one more time. Using descriptive statistics and correlation analysis, overall 23 evaluation metrics of learners are verified.

The immediate research issues are as follows. First of all, it is necessary to develop more detailed evaluation metrics depending on physical computing tools such as robots, toy, etc. Second, it is also necessary to develop evaluation standards or metrics for study materials. It is not known yet how to develop study materials in physical computing. That is, we need the general design principles and detailed guidelines.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] W. C. Jun, "Development of Evaluation Standards of Learners for Physical Computing", in 2018 International Conference on Information and Communication Technology Convergence, Jeju Island, Korea, 2018.

http://10.1109/ICTC.2018.8539466

[2] C. Kim, "A Study on the Achievement Criteria of Robot Computing Curriculum for Elementary School" Journal of the Korean Society of Information Education, **21**(1), 97-104, 2017. http://dx.doi.org/10.14352/jkaie.2017.21.1.97

[3] J. H. Kim, D. H. Kim, "Development of Physical Computing Curriculum in Elementary Schools for Computational Thinking" Journal of the Korean Society of Information Education, **20**(1), 69-82, 2016. http://dx.doi.org/10.14352/jkaie.2016.20.1.69

[4] S. Y. Choi, S. M. Kim, "Effects of Physical Computing Education Using App Inventor and Arduino on Industrial High School Students' Creative and Integrative Thinking" Journal of the Korean Association of Computer Education, **19**(6), 45-54, 2016.

[5] J. K. Shim, H. C. Kim, W. G. Lee, "A Study on Gender Differences in Programming Attitude and Achievements on Physical Computing Education in Informatics Curriculum Revised 2015" Journal of the Korean Association of Computer Education, **19**(4), 1-9, 2016.

[6] I. H. Yoo, "The Effects on Flow at Using Robots of Introductory Programming Course" Journal of the Korean Society of Information Education, **17**(3), 329-337, 2013. http://dx.doi.org/10.14352/jkaie.2013.17.3.329

[7] J. B. Song, S. H. Baik, T. W. Lee, "The Effect of Robot Programming Learning Considered Gender Differences on Female Middle School Student′s Flow Level and Problem Solving Ability" Journal of the Korean Association of Computer Education, **12**(1), 44-55, 2009..

[8] Constructivism, https://www.learning-theories.com/ constructivism.htm

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 88-96 (2019)

www.astesj.com

ASTESJ

ISSN: 2415-6698

The Effect of Fuel Additives on Pyrorated Biodiesel Blends on the Performance of a Diesel Power Generator

Maroa Semakula*, Freddie Inambao

Department of Mechanical Engineering, University of KwaZulu-Natal, Durban, South Africa

ARTICLE INFO

Article history:
Received: 14 May, 2019
Accepted: 06 August, 2019
Online: 03 September, 2019

Keywords:
2-ethyl hexyl nitrate
Ethanol
High content of oxygen
Ignition quality
Waste plastic pyrolysis oil
Phase separation

ABSTRACT

The demand for renewable energy sources such as waste pyrolysis plastic oil (WPPO), ethanol biofuel with high oxygen content is increasing globally. This study used blending of conventional diesel oil, waste plastic pyrolysis oil, ethanol and 2-ethyl hexyl nitrate (EHN) as additive. The purpose was to improve combustion, the ignition quality, and performance and emission characteristics of WPPO as an alternative source of energy. As an additive EHN reduces emissions of CO, CO_2, UHC, NO_X, and PM. On the other hand, the inclusion of ethanol purposed to improve the viscosity, increase the oxygen content of the blends, and increase miscibility of WPPO. The study utilized the following ratios, 50/WPPO25/E25, 60/WPPO20/E20, 70/WPPO15/E15, 80/WPPO10/E10 and 90/WPPO5/E5 for conventional diesel (CD), WPPO and ethanol and EHN respectively. The ratio of the additive was determined by the percentage method based on the total quantity of the blended fuel and was calculated as 0.001 %. A complete Miscibility was observed to avoid phase separation during the study and experimentation, for the blended ratios of WPPO. The testing for performance and emission characteristics was conducted on a fixed bed, water-cooled, single cylinder diesel generator engine. The results were compared to ASTM standards and discussed using tables and figures. The results conclusively show very close densities of 792 kg/m^3, 963 kg/m^3, 825 kg/m^3 for WPPO, ethanol, EHN respectively, which are close to CD fuel at 845 kg/m^3. Increased blend ratio decreased BSFC, 50 % engine load blend 80/WPPO10/E10 had values of 0.043 g/kW.h compared to full load with 0.041 g/kW.h. The highest brake thermal efficiency was by blend 90/WPPO5/E5 at 25 % engine load with a value of 26.5 % compared to 19 % at full engine load. Increase in the blend ratio and engine load decreased CO emissions up to 75 % engine load. For example, blend 90/WPPO5/E5 had a value of 0.035 % carbon emissions compared to 0.055 % at 25 % engine load. Therefore, conclusively WPPO blends can be alternative fuel with or without major engine modification.

1. Introduction

The demand for energy due to the increasing world population energy needs has placed tremendous pressure on the primary sources of energy, which are oil-based fossil fuels, in the last decade [1-3]. Although economically good on the other hand, widespread use and burning of fossil fuels is responsible for long-term environment problem in climatic changes. Secondly, fossil fuels are non-renewable and as such, their depletion is a global energy security threat. Besides depletion, the burning of fossil fuels decreases air quality leading to large-scale increase in cardiovascular diseases and cancer cases [4-6].

Globally fossil fuel still commands 80 % of the total energy requirements of the world which is consumed mostly by the transport industry alone [7]. The largest fossil fuel brand is diesel, which is a primary in diesel engines used for commercial and private transportation. Since their discovery in 1893 by Rudolph Diesel, diesel engine are efficient and good in fuel economy over power [8, 9]. Diesel engines are gaining more popularity in agricultural transport and applications, power generation and heating systems. This is due to their high thermal efficiency

*Maroa Semakula, Email: ssemakulamaroa@gmail.com

resulting into energy saving overload and safety aspects due to low volatility [10, 11].

However, diesel engines are responsible of high smoke and NOx emissions [12, 13]. Leading Diesel exhaust to be classified as carcinogenic [14] to humans which link it to increased risk of lung cancer and other cardiovascular ailments [15]. The diesel exhaust is now one of the primary sources of ground ozone [16], the sick building syndrome [17], acid rain [18], and smog [19]. Other disadvantages of fossil fuels include erratic fuel prices in the world market, potential turmoil in oil producing countries and the ever-increasing emission control requirements by governments. Therefore, finding an alternative renewable fuel that can mitigate the negative impact of fossil fuels cannot be emphasized [20].

The earliest development in alternative fuels utilized food crop-based feedstocks. As a result, this increased food insecurity in low middle-income countries and the developing countries. Hence, human rights campaigners including United Nations Food and Agriculture Organization (FAO) and the UN Security Council on human rights rejected these feedstocks. The rejection was based on Commercialization of large swathes of land for biodiesel feedstock cultivation, hence suppression of edible food crop acreage. Consequently, food prices increased and prices of essential commodities increased too [21]. Early this year a palm oil has been removed as a biodiesel feedstock by United Nations agencies like FAO due to the continued environmental hazards food security challenges in Malaysia and Indonesia.

In recent years, research has awakened focus and interest in higher alcohols due to their high levels of energy, higher cetane number (CN); better blend stability, less hygroscopic tendencies, increased carbon chain length and improved ignition quality by the alcohol fuel molecules [22]. However, lower level of alcohols methanol and ethanol, still have high qualities oxygenates. The availability of high-oxygen content with a hydroxyl (OH) group is one of the major advantages of alcohols, it reduces smoke emissions and opacity during high engine load [23, 24]. Other advantages of ethanol which make it preferable as an additive for waste plastic pyrolysis oil (WPPO) biodiesel blends include (i) ethanol is made from sources of plant origin, making it renewable, (ii) improvement in modern fermentation processes in biotechnology engineering.

This has led to improved yields in ethanol by synthesizing cellulose, for example *Clostridium species* [25], biosynthesis from glucose using genetically engineered micro-organisms like *Escherichia c*oli [26], *cyanobacteria* [27] and *Saccharomyces cerevisiae* [28]. The third reason why ethanol is important as additive in WPPO due to its inherent physicochemical composition properties. These characteristics include the presence of high oxygen content, high solubility and miscibility in WPPO blends [29-34].

Extracting liquid fuel from plastic waste material using thermal decomposition in viable and sustainable technologies has been reported in research in the last decade. Plastic waste decomposition is not only through thermal processes but also through catalytic pyrolysis. The catalytic technique uses low levels of temperature to cause plastic decomposition compared to thermal decomposition, which employs very high temperatures to produce greater yield of liquid oil. This has increased developments in turning plastic waste into energy a development that has captivated and motivated researchers such as [35-37]. A number of researchers have studied additives in their work on performance and emission impacts using biodiesel blends such as [38-53]. Especially when the final product is used with fuel additives with characteristics focusing on reducing NO_X, CO, CO_2, unburnt hydrocarbon (UHC) and particulate matter (PM) emissions [20].

A number of studies have been conducted on the reduction of NO_X and PM emissions using additives and biodiesel blends in diesel engines such as: [54] the authors used ethanol, selective catalytic reduction (SCR) catalyst Ag/Al_2O_3, and blends of biodiesel-ethanol fuel (BE). The results were mixed for emissions of UHC, CO, and PM, with an increase of 14% respectively. The authors attributed it to the presence and increase in SOFs in the PM emissions. However, a decrease in the Bosch smoke by 60 % to 80 % was reported using the European stationary cycle **(**ESC) standard. The NO_X emissions reduced by a margin of 73 % leading to a conclusion that a combination of BE and SCR catalyst arrangement provides a platform for the reduction and control of PM and NO_X emissions.

In another study by [55] wanted to determine cold flow features using ethanol-biodiesel blends by characterization. In this study a relationship between blend and torque, brake thermal efficiency, brake specific fuel consumption and emission characteristics. The findings of this study have led developed and developing countries to adopt and make fuel blending a mandatory requirement in Europe, USA and Brazil. For example, In the United States of America, the renewable fuel standard (RFS) program requires blending of fuels in increasing measure with fossil transportation fuel. This is targeting an annual growth rate escalation of 36 billion gallons by 2022 [56].

As an additive in blends, an increase in the ethanol fraction decreases auto-ignition properties of the blends, due to ethanol's low propensity [57-64]. These findings report a decrease in the CN value of the blends as the fraction of ethanol increases as the main cause. Reduced CN values prolong ignition delay and is undesirable in the combustion of the blends. Ignition delay is responsible for increased engine peak combustion pressure [65, 66]. This increases wear, combustion noises besides increasing NO_X emissions in diesel engines. The effect of CN alteration on combustion values in relation to blend characteristics is not a new research area. For example, researchers such [67-74] have covered this area and one can read the findings and conclusions.

Plastics contain a high potential of stored energy from hydrocarbon inherent in their makeup and molecular structure. Plastics are daily used in our lives and are readily available as waste in municipal solid waste sites. Plastics have become a health hazard as they litter everywhere while pausing an environmental danger to humans and animals. Therefore, altering them through modern processing methods into liquid oils for transportation fuels is a novel idea whose time has come. [75]. This study aims at utilizing and developing feedstocks from waste streams of plastics and converting them through pyrolysis into energy. This will provide a framework for creating alternative fuels to reduce dependency on petroleum fuels, which are primary transport fuels. The third aspect of this work is to show the importance of blending WPPO or any other new alternative feedstocks to improve their

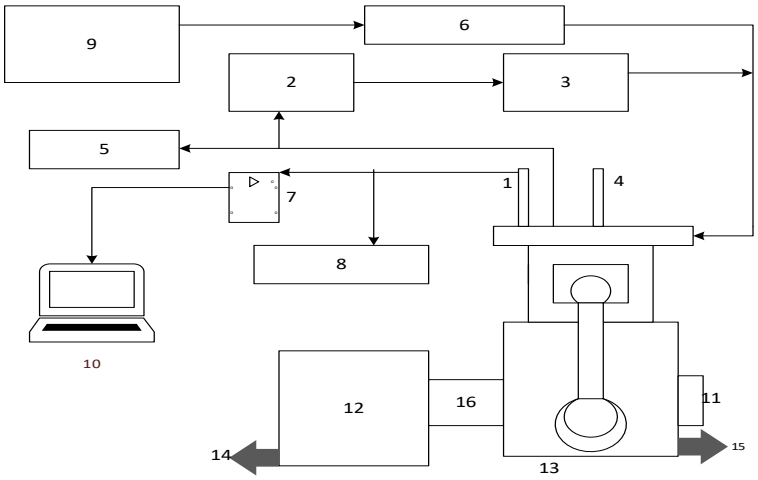


Figure 1: Schematic diagram of the test engine rig

Key: 1 Cylinder pressure sensor; 2 EGR control valve; 3 EGR cooler; 4 Injection control unit; 5 Exhaust gas exit; 6 Air box; 7 Signal amplifier; 8 Gas analyser; 9 Air flow meter; 10 Data acquisition system; 11 Crank position sensor; 12 Dynamometer; 13 Engine; 14 Cooling water exit from the dynamometer to the cooling tower; 15 Cooling water exit from engine to the cooling tower; 16 Dynamometer drive coupling

qualities. The focus is to reduce emission and enhance performance of transportation fuel while creating jobs in recycling but protecting our environment.

2. Methodology

2.1. Methodology and Experimental Set-Up

This experiment is making a case for blending of WPPO whose n-alkenes are very low by 25 % in auto-ignition, compared to diesel fuel whose n-alkenes are good for auto-ignition. The aromatics, which affect PM emissions, are very low in WPPO blends. WPPO consists of iso-alkanes, n-alkanes, and olefins in the areas of 27 %, 25 %, and 9 % respectively, the remaining 30 % content is undefined due to complicated chemical bond structures [76, 77]. However, aromatics cyclo-alkanes (naphthalene) and other poor in auto-ignition were also found to be 40 % by [78]. Blending was preferred to improve the low pour point of WPPO to improve its cold starting characteristics. Secondly blending improves the fuel spray characteristics. Blending using ethanol, which is soluble and miscible in WPPO blends, also improves physicochemical properties of blends. Thirdly blending helped this experiment to reduce the viscosity of WPPO biodiesel, thus aiding and improving spray characteristics. The blends used in this experiment are broken into five ratios namely: 50/WPPO25/E25, 60/WPPO20/E20, 70/WPPO15/E15, 80/WPPO10/E10 and 90/WPPO5/E5 respectively. For example, to define a blend sample such as 90/WPPO5/E5 is composed of 90 % conventional diesel (CD) fuel by volume 5 % WPPO and 5 % additive 2-Ethyhexyl Nitrate (EHN) respectively for all the replicates.

2.2. Engine Tests

The experiment used a naturally aspirated single-cylinder diesel engine power generator, water cooled, direct injection, Kirloskar TV1, in the Mechanical Engineering Department Laboratory, the University of KwaZulu Natal in Durban, South Africa. The details of the engine and specifications are in Table 1. Figure 1 shows a schematic of the engine test setup.

Table 1: Experimental engine specification

Parameters	Position value
Ignition Type	4 (Stroke)DICI
Number Of Cylinders	1
Model	TV 1
Cooling Medium	Water
Manufacturer	Kirloskar
Revolutions Per Minute	1500
Brake Power	3.5 kW
Cylinder Bore	87.5 mm
Piston Stroke	110 mm
Compression Ratio	18.5:1
Connecting-Rod Length	234
Engine Capacity	661cc
Dynamometer Make	234
Injection Timing	23.4˚ bTDC
Maximum Torque	28 Nm @1500
Injection Pressure	250 Bar

2.3. Physicochemical Property Analysis

The WPPO pyrolysis unit was designed at the University of KwaZulu Natal School of mechanical engineering GPS (29°52’09.9”S 30°58’37.9”E), by the author who is from the Green Energy Group, Howard campus, Durban, South Africa. Lucien engineering Company fabricated the final product in their premises in Durban. The pyrolysis unit comprised of the heating system and its control, the reactor tank, pipes, the heat exchanger and the collector, which all culminate into the pyrolysis unit. The energy requirements for each step is according to the references [79, 80], and the plant’s production chart is shown in Figure 3. Ethanol, conventional diesel and EHN in local outlets and blended using a homogenizer for 5 min at 3000 rpm. Sampling and testing for properties was in the Department of Chemical Engineering Laboratory, at the University of KwaZulu Natal in Durban, South Africa. Table 2 shows some important physicochemical properties of the fuels oils before blending.

Table 2: Properties of Diesel, WPPO, and ethanol before blending and the addition of EHN

Properties	Unit	CD	WPPO	Ethanol
Density @ 20ºC	kg/M^3	845	825	792
Visc.@ 40ºC	cSt	3.04	2.538	1.05
Cetane Number		55		8.5
Flash Point	º C	50	43	16
Fire Point	º C	56	45	53
Carbon residue	%	22	0.015	
Sulfur content	%	<0.028	0.248	
Gross Calories	kJ/kg	46500	43340	29700
Cetane index		46	65	

Table 3: Properties of blended ratio mixtures of diesel, ethanol, WPPO with EHN

Property	Unit	CD	90/5/5	80/10/10	70/15/15	60/20/20	50/25/25	STANDARD
Density	Kg/M^3	845	838.5	834	830	825	823	ASTM D1298
KViscosity@40	cST	3.452	2.38	2.37	2.365	2.340	2.325	ASTM D445
Cetane Number	-	48	58	60	62	63	65	ASTM D4737
GCV	kJ/kg	44840	40125	39985	38700	36800	34500	ASTM D4868
Sulfur Content	%	<0.0124	0.0248	0.0249	0.0251	0.0253	0.0257	ASTM D4294
Oxygen	%	12.35	11.80	10.75	10.15	10.25	10.35	ASTM D5622
Carbon Residue	%	74.85	75.35	76.40	77.55	78.25	79.65	ASTM D7662
Flashpoint	ºC	56.5	38.5	37.55	37.35	37.15	36.85	ASTM D93
Hydrogen	%	12.38	7.5	7.55	7.65	7.75	7.95	ASTM D7171

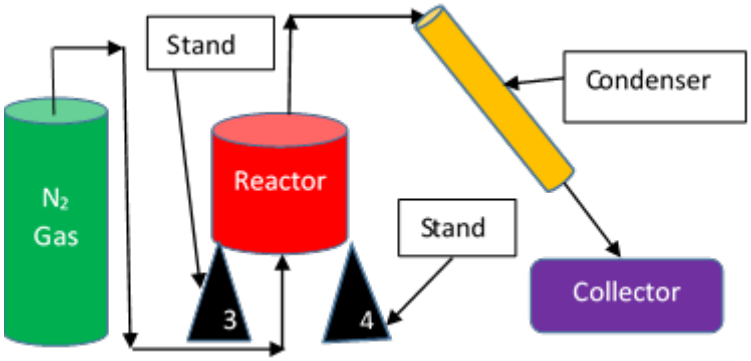


Figure 2: Assembled view of the pyrolysis unit experimental set-up

Table 3 shows the physicochemical properties of blended mixtures fuels and their determined fuel properties after blending in relation to American Standard of Testing and Measurements (ASTM).

3. Results and Discussion

3.1. Brake Specific Fuel Consumption

Figure 3 is a variation of brake specific fuel consumption (BSFC) with engine load. The BSFC compared to the engine load in Fig. 3 shows that as the load increases there is an equal increase for fuel consumed by the test engine. The values obtained at full engine load for the blends of 90/WPPO5/E5, 80/WPPO10/E10, 70/WPPO15/E15, 60/WPPO20/E20, 50/WPPO25/E25 and CD were 0.04 g/kW.h, 0.041 g/kW.h, 0.042 g/kW.h, 0.043 g/kW.h and 0.035 g/kW.h respectively.

At high engine loads the conversion of heat energy to mechanical energy increases with increase in combustion temperature, leading to increased BSFC for the biodiesel. This increase is proportional to the difference in their heating values which is identical to the findings of [81]. Additionally, WPPO blends have high densities, therefore suffer high mass injection pressure, hence the increase in BSFC which is identical to studies by [82, 83]. These blends of WPPO compare well to conventional diesel fuel and sometimes-other biodiesel blends with comparative differences in the heating values.

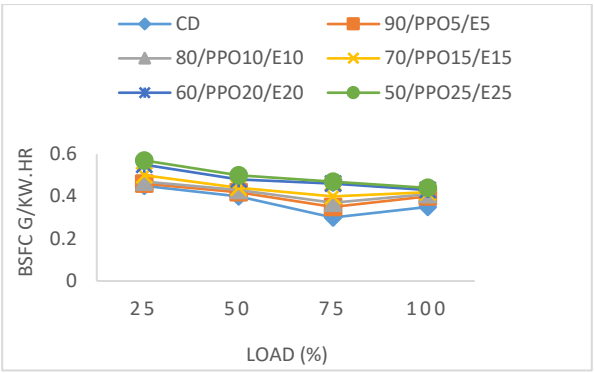


Figure 3: Brake specific fuel consumption versus load

As the blend ratio increased there was a decrease in the BSFC across all the test fuels. However, the values for all WPPO blends increased compared to CD test fuel. This is due to the lower calorific values of the blends as the percentage of the blend ratio

increased. In other words, by increasing the ratio of WPPO in the diesel test fuel, the engine fuel consumption increased, this is identical to the studies of [84-86]. The closeness of the values and the packed graph reveal a close resemblance and identical BSFC characteristics of WPPO, ethanol, and EHN compared to CD fuel. For example, at 50 % engine load the blend of 80/WPPO10/E10 had a value of 0.043 g/kW.h compared to full engine load with 0.041 g/kW.h. This value is higher than CD test fuel with 0.04g/kW.h at 50 % engine load and 0.035 g/kW.h at full engine load.

3.2. Brake Thermal Efficiency

The brake thermal efficiency (BTE) variations with engine load is as shown in Figure 4. The graphs show that, as the load increased there was an increase in the BTE across all the test fuel blends of WPPO and CD. The result of this experiment shows that the BTE increased as the load increased, explained by the reduction in the heat loss as the engine power (more fuel) increased with load. At 50 % engine load the values for blends 90/WPPO5/E5, 80/WPPO10/E10, 70/WPPO15/E15, 60/WPPO20/E20, 50/WPPO25/E25 and CD were 22 %, 21 %, 20 %, 18 %, 16.5 % and 22.5 % respectively. As the blend ratio and engine load increased, there was an increase in BTE across the blends of WPPO but with a decrease in the BTE within the blends. For example, at 25 % engine load, 90/WPPO5/E5 had values of 14 %, 22 %, 26.5 % and 25 % compared to 70/WPPO15/E15 with 12.5 %, 20 %, 22.5 % and 23 % respectively.

The highest BTE value was by blend 90/WPPO5/E5 at 25 % engine load compared to any other blend of WPPO. For example, Fig. 4 shows values of 24.8 %, 23 %, 21 % and 19 % respectively for blends 80/WPPO10/E10, 70/WPPO15/E15, 60/WPPO20/E20, 50/WPPO25/E25. However, blend 50/WPPO25/E25 reported the lowest values compared to the other blends. At 25 % engine load the BTE value was 9.5 % compared with 19 % at full load, these two are the lowest values of BTE.

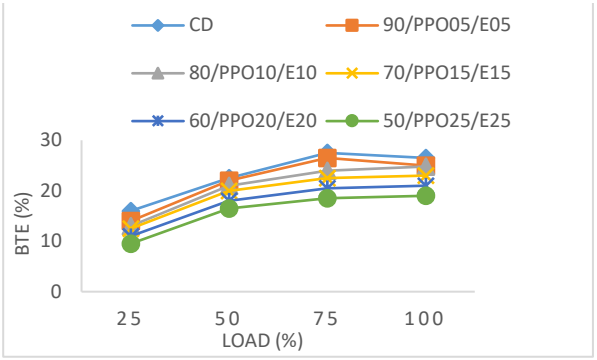


Figure 4: Brake thermal efficiency versus load

3.3. Exhaust Gas Temperature

Figure 5 is a variation of exhaust gas temperature (EGT) and engine load. The graph reveals that as the load increases the value of the EGT increased significantly and linearly especially for the blends. At 25 % engine load the blends 90/WPPO5/E5, 80/WPPO10/E10, 70/WPPO15/E15, 60/WPPO20/E20; 50/WPPO25/E25 reported values of 165 ºC, 195 ºC, 226 ºC and 256 ºC compared to CD with 155 ºC, 175 ºC, 205 ºC and 240 ºC for all engine load conditions. What explains this increase in engine temperatures is the increase in the required fuel to compensate increased load. However, observed also is temperature increase becomes pronounced with increased blend ratio of WPPO compared to CD test fuel as in Figure 5. The second reason for increased EGT temperature increase is the BTE heat loss factor for high blend ratios compared to CD test fuel.

The experiment also showed another interesting observation, as the engine load increased from 25 % to full load (100 %) the graph curves tend toward unitary, adopting almost identical and similar values to CD test fuel. This concludes that the blends of WPPO, ethanol, and fuel additives have identical temperature characteristics to those of CD test fuel especially as the engine load hits 75 % heading to 100 % (full load). This is due to the presence of ethanol, which decreased ignition delay thus lowering the combustion temperature. The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multi-levelled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

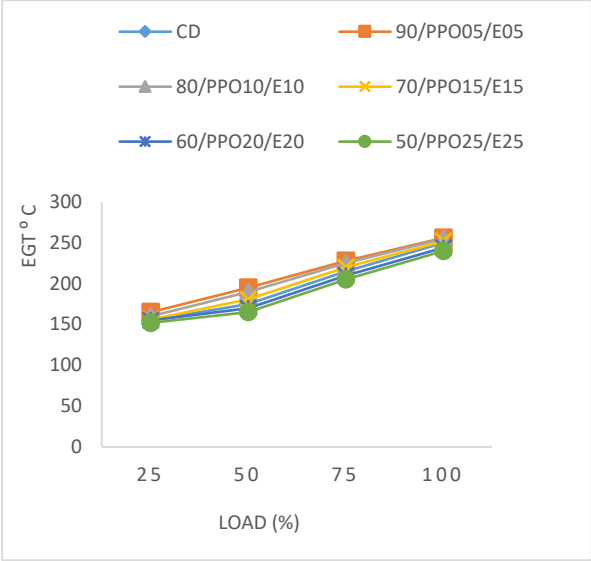


Figure 5: Exhaust gas temperature versus load

3.4. Carbon Monoxide

Figure 6 is a variation of carbon monoxide (CO) with engine load. The graph reveals that as the engine load and the blend ratio increased CO emissions decreased up to 75 % of engine load. This was for all blends 90/WPPO5/E5, 80/WPPO10/E10, 70/WPPO15/E15, 60/WPPO20/E20, and 50/WPPO25/E25. The highest value in Figure 6 of CO emission reported was 0.0625 % for blend 50/WPPO25/E25 and the lowest value reported was by blend 90/WPPO5/E5 at 0.055 %. Another observation is that as the engine was approaching full load, all the test fuels showed increased CO emissions with blends 90/WPPO5/E5 and 80/WPPO10/E10 reporting the lowest emissions among the test blends across all the engine load conditions.

For example, at 25 % engine load the blends of 90/WPPO5/E5, 80/WPPO10/E10, 70/WPPO15/E15, 60/WPPO20/E20, 50/WPPO25/E25 reported values of 0.055 %, 0.0565 %, 0.06 %, 0.0615 % and 0.0625 %. However, as the load increased for

example, from 25 % to 75 % the values reported are 0.035 %, 0.0375 %, 0.0445 % and 0.0.0475 respectively.

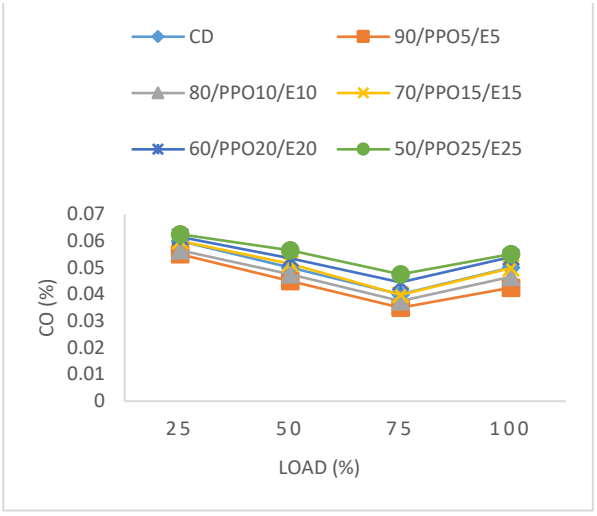


Figure 6: Carbon monoxide versus load

There are a number of factors, which explain the decrease in CO emissions as the engine load is increasing. The decrease can be attributed to increased unburnt or partially burnt hydrocarbons due to incomplete combustion, which is identical to the study by [87]. This phenomenon is linked to the increased load and the presence of ethanol, which shortened ignition delay, hence increasing CO emissions. Additionally, the decrease in CO emissions could also be due to the conversion of CO to CO_2 taking up this reaction from the high oxygen content of the fuel additive ethanol.

3.5. Carbon Dioxide

Figure 7 is the variation of carbon dioxide (CO_2) with engine load. The graph shows that as the blend ratio and engine load increased CO_2 emissions increased but compared to CD their emission levels are still lower and almost identical. At 50 % engine load the values of CD, and the blends of 90/WPPO5/E5, 80/WPPO10/E10, 70/WPPO15/E15, 60/WPPO20/E20, 50/WPPO25/E25 were 3.58 %, 3.35 %, 2.95 %, 2.6 %, 2.55 % and 2.25 % respectively.

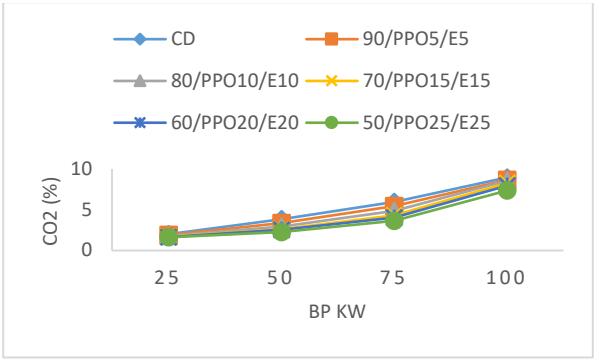


Figure 7: Carbon dioxide versus load

3.6. Oxides of Nitrogen

Figure 8 is a variation of engine load with oxides of nitrogen (NO_X) emissions. The graph plot shows that as the engine load was increased there was an increase in the NO_X emissions irrespective of fuel, blend ratio or EHN. However, the value of NO_X emissions from the blends 90/WPPO5/E5, 80/WPPO10/E10, and 70/WPPO15/E15 reported lower values compared to CD fuel. For example, at 50 % the values of the blends, was 385 ppm, 396 ppm, and 415 ppm, compared to CD fuel at 425 ppm.

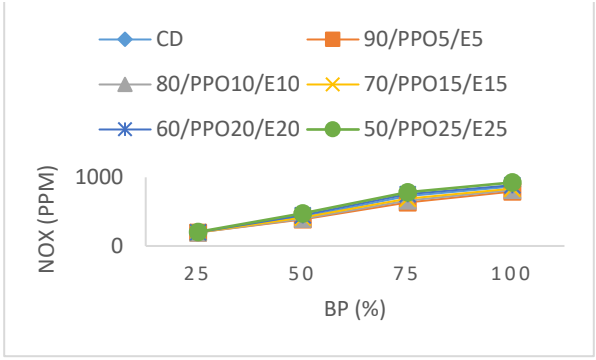


Figure 8: Oxides of nitrogen versus load

Blend 60/WPPO20/E20 and 50/WPPO25/E25 had the highest NO_X emissions compared to the other blends of 90/WPPO5/E5, 80/WPPO10/E10, and 70/WPPO15/E15 across all the engine load conditions tested. In Figure 8 at 25 % engine load, the two blends had values of 205 ppm and 200 ppm respectively. However, at full engine load the NO_X emissions values increased to 925 ppm and 885 ppm compared to blend 90/WPPO5/E5 at the same load with 197 ppm and at full load at 792 ppm.

From the graph plot in Figure 8, as the blend ratio increased there was a direct increase in emissions of NO_X across all the blended test fuels. However, blend 90/WPPO5/E5 reported the lowest values of NO_X emissions compared to all the other tested blends experimented. The formation of NO_X in biodiesel fuel combustion strongly depends on the combustion temperatures and the oxygen concentration in the combustion zone. The low blend ratios of 70/WPPO15/E15, 60/WPPO20/E20, 50/WPPO25/E25, show a shortened combustion process. Hence, a poor cooling effect and failure to decrease peak combustion temperatures leading to increased NO_X.

These findings show a correlation between the alcohol content in the fuel and peak flame temperatures, the content of nitrogen, and oxygen availability. this study findings are identical to the findings of [88]. The increased NO_X emissions are a result of the presence of nitrogen from the CN improver ENH and other contaminants from the WPPO impurities. Additionally, it could be due to the generation of radicals of hydrocarbon through molecular unsaturation in the blends being identical to the findings of [89, 90]. The final factor is due to increased EGT temperatures, linked to the high oxygen content and the air fuel ratio.

3.7. Unburnt Hydrocarbons

Figure 9 is a variation of unburnt hydrocarbons (UHC) emission with engine load. As the engine load was increased, the UHC emissions increased too. The higher hydrocarbon emissions may be due to hydrogen radicals in the diesel-ethanol-WPPO-EHN blends. The high fraction of ethanol in blends 70/WPPO15/E15, 60/WPPO20/E20, 50/WPPO25/E25 contributes to increase in the emissions of UHC which is identical to the findings of [62, 91]. Who observed it in an SI engine cylinder walls, crevices, and quenched cylinder walls with richer air-alcohol mixtures.

However, the increase is more significant as the engine load was in intermediate loads of 75 % moving to or approaching full

load. For example, at 50 % engine load, the values of blends were 22 ppm, 21 ppm, 20 ppm, 18 ppm, and 15 ppm respectively compared to full load with 35 ppm, 34 ppm, 32 ppm, 29 ppm, and 26 ppm (for blends 90/WPPO5/E5, 80/WPPO10/E10, 70/WPPO15/E15, 60/WPPO20/E20, and 50/WPPO25/E25). This leads to the conclusion that at high engine loads the values of UHC emissions are significantly high for all the blends of WPPO, ethanol and EHN, although comparatively low compared to CD fuel.

The UHC emissions from the blends 90/WPPO5/E5 and 80/WPPO10/E10 report higher values. Although, the graph plot in Figure 9 shows low values compared to the values of CD test fuel. However, the general trend by the graph in Figure 9 shows that increased blend ratio significantly reduced UHC emissions, across all the test fuels irrespective of the engine load condition. This reduction is due to the high oxygen content of ethanol and the effect of the fuel additive EHN, factors that helped in complete combustion.

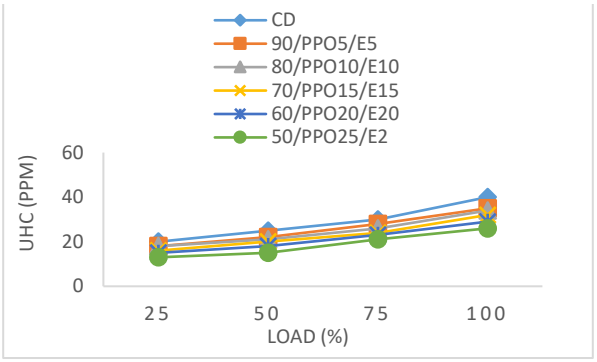


Figure 9: Unburnt hydrocarbons versus load

4. Conclusion

- The variation of BSFC with engine load shows that, as the load increases there is an increase in the fuel consumed by the test engine as in the graph as shown in Fig.1. However, the lower blend ratios 90/WPPO5/E5 and 80/WPPO10/E10 exhibit identical BSFC values to conventional diesel test fuel compared to the other blends. These blends show the lowest BSFC values compared to the other WPPO blends.
- The BTE of blend 90/WPPO5/E5 report values close to the values of conventional diesel fuel values. This is due to the close blend density values and the gross calorific values of WPPO blends, which showed very small and marginal differences. This is the case condition at lower blend ratios for WPPO blends tested.
- There is a reduction of UHC emissions with the use of WPPO blends, ethanol and EHN, with a notable reduction in NOX emissions, especially for the blend 90/WPPO5/E. This is a clear indication that this blend performed well when compared to petroleum conventional diesel.
- Although there is an indicated increase in the emissions of CO, CO_2 NOX and UHC, for all the blends of WPPO, ethanol, and EHN. There is a clear indication in the graph plots discussed that blended fuels have low emission levels compared to conventional petroleum diesel in relation to the ASTM standards during experimentation.
- During experimentation, the blends of WPPO, ethanol, and EHN report identical temperature characteristics to CD test fuel as the engine load hits 75 % heading to 100 % (full load). This is due to the presence of ethanol responsible for decreased ignition delay.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The writer would lie to thank the Green Energy Group for their equipment and material support in terms of logistics and transport during the course of the study

References

[1] A. Fayyazbakhsh, V. Pirouzfar, “Investigating the influence of additives-fuel on diesel engine performance and emissions: Analytical modeling and experimental validation” Fuel, 171, 167–177, 2016. https://doi.org/10.1016/j.fuel.2015.12.028

[2] A. Fayyazbakhsh, V. Pirouzfar, “Determining the optimum conditions for modified diesel fuel combustion considering its emission, properties and engine performance” Energy Convers. Manage., 113, 209–219, 2016. https://doi.org/10.1016/j.enconman.2016.01.058

[3] S. Khorramshokouh, V. Pirouzfar, Y. Kazerouni, A. Fayyazbakhsh, R. Abedini, “Improving the properties and engine performance of diesel–methanol–nanoparticle blend fuels via optimization of the emissions and engine performance” Energy Fuels, 30(10), 8200–8208, 2016. https://doi.org/10.1021/acs.energyfuels.6b01856

[4] V. Pirouzfar, A. Fayyazbakhsh, Diesel Fuel Additives, LAP LAMBERT Academic Publishing, Riga, Latvia, 2016.

[5] A. K. Agarwal, “Biofuels (alcohols and biodiesel) applications as fuels for internal combustion engines” Prog. Energy Combust. Sci., 33(3), 233–271, 2007. https://doi.org/10.1016/j.pecs.2006.08.003

[6] C. L. Song, Y. C. Zhou, R. J. Huang, Y. Q. Wang, Q. F. Huang, G. Lü, K. M. Liu, “Influence of ethanol–diesel blended fuels on diesel exhaust emissions and mutagenic and genotoxic activities of particulate extracts” J. Hazard. Mater., 149(2), 355–363, 2007. https://doi.org/10.1016/j.jhazmat.2007.03.088

[7] J. C. Escobar, E. S. Lora, O. J. Venturini, E. E. Yáñez, E. F. Castillo, O. Almazan, “Biofuels: environment, technology and food security” Renewable Sustainable Energy Rev., 13(6-7), 1275–1287, 2009. https://doi.org/10.1016/j.rser.2008.08.014

[8] S. Maroa, F. Inambao, “The formation, effects and control of oxides of nitrogen in diesel engines” Int. J. Eng. Res., 13(6), 3200–3209, 2018.

[9] S. Maroa, F. Inambao, “Transportation, pollution and the environment” Int. J. Eng. Res., 13(6), 3187–3199, 2018.

[10] J. B. Heywood, Internal Combustion Engine Fundamentals, McGraw-Hill Education, New Delhi, 2012.

[11] A. M. Taylor, “Science review of internal combustion engines” Energy Policy, 36(12), 4657–4667, 2008. https://doi.org/10.1016/j.enpol.2008.09.001

[12] M. Fu, Y. Ge, J. Tan, T. Zeng, B. Liang, “Characteristics of typical non-road machinery emissions in China by using portable emission measurement system” Sci. Total Environ., 437, 255–261, 2012. https://doi.org/10.1016/j.scitotenv.2012.07.095

[13] Y. Zhang, X. Wang, G. Li, W. Yang, Z. Huang, Z. Zhang, X. Huang, D. Wei, T. Liu, Z. Huang, et al., “Emission factors of fine particles, carbonaceous aerosols and traces gases from road vehicles: recent tests in an urban tunnel in the Pearl River Delta, China” Atmos. Environ., 122, 876–884, 2015. https://doi.org/10.1016/j.atmosenv.2015.08.024

[14] L. Benbrahim-Tallaa, R. A. Baan, Y. Grosse, B. Lauby-Secretan, F. El Ghissassi, V. Bouvard, N. Guha, D. Loomis, K. Straif, “Carcinogenicity of diesel-engine and gasoline-engine exhausts and some nitroarenes” Lancet Oncol., 13(7), 663–664, 2012. https://doi.org/10.1016/S1470-2045(12)70280-2

[15] L. V. Giles, C. Carlsten, M. S. Koehle, “The effect of pre-exercise diesel exhaust exposure on cycling performance and cardio-respiratory variables” Inhal. Toxicol., 24(12), 783–789, 2012. DOI: 10.3109/08958378.2012.717649

[16] W. B. Innes, “Effect of nitrogen oxide emissions on ozone levels in metropolitan regions” Environ. Sci. Technol., 15(8), 904–912, 1981. https://doi.org/10.1021/es00090a003

[17] R. E. Hester, R. M. Harrison (eds), Air Quality in Urban Environments. Royal Society of Chemistry, 2009. https://doi.org/10.1039/9781847559654
[18] A. Mills, S. Elouali, "The nitric oxide ISO photocatalytic reactor system: measurement of NOx removal activity and capacity" J. Photochem. Photobiol., A, 305, 29–36, 2015. https://doi.org/10.1016/j.jphotochem.2015.03.002
[19] J. Ou, Z. Yuan, J. Zheng, Z. Huang, M. Shao, Z. Li, X. Huang, H. Guo, P. K. K. Louie, "Ambient ozone control in a photochemically active region: short-term despiking or long-term attainment?" Environ. Sci. Technol., 50(11), 5720–5728, 2016. https://doi.org/10.1021/acs.est.6b00345
[20] D. Damodharan, A. P. Sathiyagnanam, B. Rajesh Kumar, K. C. Ganesh, "Cleaner emissions from a DI diesel engine fueled with waste plastic oil derived from municipal solid waste under the influence of n-pentanol addition, cold EGR, and injection timing" Environ. Sci. Pollut. Res., 25(14), 13611–13625, 2018. https://doi.org/10.1007/s11356-018-1558-5
[21] B. R. Kumar, S. Saravanan, "Use of higher alcohol biofuels in diesel engines: A review" Renewable Sustainable Energy Rev., 60, 84–115, 2016. https://doi.org/10.1016/j.rser.2016.01.085
[22] E. Koivisto, N. Ladommatos, M. Gold, "Systematic study of the effect of the hydroxyl functional group in alcohol molecules on compression ignition and exhaust gas emissions" Fuel, 153, 650–663, 2015. https://doi.org/10.1016/j.fuel.2015.03.042
[23] M. Lapuerta, R. García-Contreras, J. Campos-Fernández, M. Pilar Dorado, "Stability, lubricity, viscosity, and cold-flow properties of alcohol-diesel blends" Energy Fuels, 24(8), 4497–4502, 2010. https://doi.org/10.1021/ef100498u
[24] Y. Ren, Z. Huang, H. Miao. Y. Di, D. Jiang, K. Zeng, B. Liu, X. Wang, "Combustion and emissions of a DI diesel engine fueled with diesel-oxygenate blends" Fuel, 87(12), 2691–2697, 2008. https://doi.org/10.1016/j.fuel.2008.02.017
[25] S.M. Gaida, A. Liedtke, A. H. W. Jentges, B. Engels, S. Jennewein, "Metabolic engineering of Clostridium cellulolyticum for the production of n-butanol from crystalline cellulose" Microbial Cell Factories, 15(1), 6, 2016. https://doi.org/10.1186/s12934-015-0406-2
[26] S. H. Desai, C. A. Rabinovitch-Deere, Z. Fan, S. Atsumi, "Isobutanol production from cellobionic acid in Escherichia coli" Microbial Cell Factories, 14(1), 52, 2015. https://doi.org/10.1186/s12934-015-0232-6
[27] C. Formighieri, Solar-to-Fuel Conversion in Algae and Cyanobacteria, Springer, 2015, 31–38.
[28] E. Ofuonye, K. Kutin, D. T. Stuart, "Engineering Saccharomyces cerevisiae fermentative pathways for the production of isobutanol" Biofuels, 4(2), 185–201, 2013. https://doi.org/10.4155/bfs.12.85
[29] S. Fernando, M. Hanna, "Development of a novel biofuel blend using ethanol-biodiesel-diesel microemulsions: EB-diesel" Energy Fuels, 18(6), 1695–1703, 2004. DOI:10.1021/ef049865e
[30] P. Kwanchareon, A. Luengnaruemitchai, S. Jai-In, "Solubility of a diesel–biodiesel–ethanol blend, its fuel properties, and its emission characteristics from diesel engine" Fuel, 86(7-8), 1053–1061, 2007. https://doi.org/10.1016/j.fuel.2006.09.034
[31] M. Lapuerta, O. Armas, R. Garcia-Contreras, "Stability of diesel–bioethanol blends for use in diesel engines" Fuel, 86(10-11), 1351–1357, 2007. https://doi.org/10.1016/j.fuel.2006.11.042
[32] M. Lapuerta, O. Armas, J. Rodriguez-Fernandez, "Effect of biodiesel fuels on diesel engine emissions" Prog. Energy Combust. Sci., 34(2), 198–223, 2008. https://doi.org/10.1016/j.pecs.2007.07.001
[33] D.-G. Li, H. Zhen, L. Xingcai, Z. Wu-Gao, Y. Jian-Guang, "Physico-chemical properties of ethanol–diesel blend fuel and its effect on performance and emissions of diesel engines" Renewable Energy, 30(6), 967–976, 2005. https://doi.org/10.1016/j.renene.2004.07.010
[34] S. Shahir, H. H. Masjuki, M. A. Kalam, A. Imran, A. M. Ashraful, "Performance and emission assessment of diesel–biodiesel–ethanol/bioethanol blend as a fuel in diesel engines: A review" Renewable Sustainable Energy Rev., 48, 62–78, 2015. https://doi.org/10.1016/j.rser.2015.03.049
[35] K. Zhang, M. R. Sawaya, D. S. Eisenberg, J. C. Liao, "Expanding metabolism for biosynthesis of nonnatural alcohols" Proc. Natl. Acad. Sci. 105(52), 20653–20658, 2008. https://doi.org/10.1073/pnas.0807157106
[36] A. F. Cann, J. C. Liao, "Pentanol isomer synthesis in engineered microorganisms" Appl. Microbiol. Biotechnol. 85(4), 893–899, 2010. doi: 10.1007/s00253-009-2262-7.
[37] Y. Dekishima, E. I. Lan, C. R. Shen, K. M. Cho, J. C. Liao, "Extending carbon chain length of 1-butanol pathway for 1-hexanol synthesis from glucose by engineered Escherichia coli" J. Am. Chem. Soc., 133(30), 11399–11401, 2011. DOI: 10.1021/ja203814d
[38] J. Campos-Fernández, J. M. Arnal, J. Gómez, M. Pilar Doradoa, "A comparison of performance of higher alcohols/diesel fuel blends in a diesel engine" Appl. Energy, 95, 267–275, 2012. https://doi.org/10.1016/j.apenergy.2012.02.051
[39] Z. Chen, J. Liu, Z. Han, B. Du, Y. Liu, C. Lee, "Study on performance and emissions of a passenger-car diesel engine fueled with butanol-diesel blends" Energy, 55, 638–646, 2013. https://doi.org/10.1016/j.energy.2013.03.054
[40] Z. Chen, J. Liu, Z. Wu, C. Lee, "Effects of port fuel injection (PFI) of n-butanol and EGR on combustion and emissions of a direct injection diesel engine" Energy Convers. Manage., 76, 725–731, 2013. https://doi.org/10.1016/j.enconman.2013.08.030
[41] J. Devaraj, Y. Robinson, P. Ganapathi, "Experimental investigation of performance, emission and combustion characteristics of waste plastic pyrolysis oil blended with diethyl ether used as fuel for diesel engine" Energy, 85, 304–309, 2015. https://doi.org/10.1016/j.energy.2015.03.075
[42] C. Jin, M. Yao, H. Liu, C. F. Lee, J. Ji, "Progress in the production and application of n-butanol as a biofuel" Renewable Sustainable Energy Rev., 15(8), 4080–4106, 2011. https://doi.org/10.1016/j.rser.2011.06.001
[43] B. R. Kumar, S. Saravanan, "Effects of iso-butanol/diesel and n-pentanol/diesel blends on performance and emissions of a DI diesel engine under premixed LTC (low temperature combustion) mode" Fuel, 170, 49–59, 2016. DOI: 10.1016/j.fuel.2015.12.029
[44] S. Kumar, R. Prakash, S. Murugan, R. K. Singh, "Performance and emission analysis of blends of waste plastic oil obtained by catalytic pyrolysis of waste HDPE with diesel in a CI engine" Energy Convers. Manage., 74, 323–331, 2013. https://doi.org/10.1016/j.enconman.2013.05.028
[45] M. Mani, G. Nagarajan, S. Sampath, "Characterization and effect of using waste plastic oil and diesel fuel blends in compression ignition engine" Fuel Energy Abstracts, 36(1), 212–219, 2011. https://doi.org/10.1016/j.energy.2010.10.049
[46] S. H. Park, J. Cha, H. J. Kim, C. S. Lee, "Effect of early injection strategy on spray atomization and emission reduction characteristics in bioethanol blended diesel fueled engine" Energy, 39(1), 375–387, 2012. https://doi.org/10.1016/j.energy.2011.12.050
[47] C. D. Rakopoulos, A. M. Dimaratos, E. G. Giakoumis, D. C. Rakopoulos, "Investigating the emissions during acceleration of a turbocharged diesel engine operating with bio-diesel or n-butanol diesel fuel blends" Energy, 35(12), 5173–5184, 2010. https://doi.org/10.1016/j.energy.2010.07.049
[48] S. Saravanan, "Effect of exhaust gas recirculation (EGR) on performance and emissions of a constant speed DI diesel engine fueled with pentanol/diesel blends" Fuel, 160, 217–226, 2015. https://doi.org/10.1016/j.fuel.2015.07.089
[49] V. Soloiu, M. Duggan, S. Harp, B. Vlcek, D. Williams, "PFI (port fuel injection) of n-butanol and direct injection of biodiesel to attain LTC (low-temperature combustion) for low-emissions idling in a compression engine" Energy, 52, 143–154, 2013. https://doi.org/10.1016/j.energy.2013.01.023
[50] X. Wang, C. S.Cheung, Y. Di, Z. Huang, "Diesel engine gaseous and particle emissions fueled with diesel-oxygenate blends" Fuel, 94, 317–323, 2012. https://doi.org/10.1016/j.fuel.2011.09.016
[51] L. Wei, C. Cheung, Z. Huang, "Effect of n-pentanol addition on the combustion, performance and emission characteristics of a direct-injection diesel engine" Energy, 70, 172–180, 2014. https://doi.org/10.1016/j.energy.2014.03.106
[52] L. Zhu, C. S. Cheung, W. G. Zhang, Z. Huang, "Combustion, performance and emission characteristics of a DI diesel engine fueled with ethanol–biodiesel blends" Fuel, 90(5), 1743–1750, 2011. https://doi.org/10.1016/j.fuel.2011.01.024
[53] Q. Zhang, M. Yao, Z. Zheng, H. Liu, J. Xu, "Experimental study of n-butanol addition on performance and emissions with diesel low temperature combustion" Energy, 47(1), 515–521, 2012. https://doi.org/10.1016/j.energy.2012.09.020
[54] S. Xiaoyan, Y. Yunbo, H. Hong, S. Shijin, D. Hongyi, L. Rulong, "Combination of biodiesel-ethanol-diesel fuel blend and SCR catalyst assembly to reduce emissions from a heavy-duty diesel engine" J. Environ. Sci., 20(2), 177–182, 2008. https://doi.org/10.1016/S1001-0742(08)60028-5
[55] F. Aydın, H. Öğüt, "Effects of using ethanol-biodiesel-diesel fuel in single cylinder diesel engine to engine performance and emissions" Renewable Energy, 103, 688–694, 2017. https://doi.org/10.1016/j.renene.2016.10.083
[56] K. Lawyer, A. Ickes, T. Wallner, D. Ertl, R. Williamson, S. Miers, J. Naber, "Blend Ratio Optimization of Fuels Containing Gasoline Blendstock, Ethanol, and Higher Alcohols (C3-C6): Part I-Methodology and Scenario Definition" 2013, SAE Technical Paper 2013-01-1144. https://doi.org/10.4271/2013-01-1144
[57] Ö. Can, I. Celikten, N. Usta, "Effects of ethanol addition on performance and emissions of a turbocharged indirect injection diesel engine running at different injection pressures" Energy Convers. Manage., 45(15–16), 2429-2440, 2004. https://doi.org/10.1016/j.enconman.2003.11.024
[58] C. Chacartegui, J. Lopez, F. Alfonso, P. Aakko, C. Hamelinck, G. Vossen, H. Kattenwinkel, "Blending ethanol in diesel" Final report for Lot 3b of the Biodiesel Improvement on Standards. Coordination of Producers and Ethanol Studies (Bioscopes) project, 2007.
[59] A. C. Hansen, Q. Zhang, P. W. Lyne, "Ethanol-diesel fuel blends—a review", Bioresour. Technol., 96(3), 277–285, 2005. https://doi.org/10.1016/j.biortech.2004.04.007

[60] H. Kim, B. Choi, "Effect of ethanol–diesel blend fuels on emission and particle size distribution in a common-rail direct injection diesel engine with warm-up catalytic converter" Renewable Energy, 33(10), 2222–2228, 2008. https://doi.org/10.1016/j.renene.2008.01.002
[61] H. Kuszewski, A. Jaworski, A. Ustrzycki, "Lubricity of ethanol–diesel blends–Study with the HFRR method" Fuel, 208, 491–498, 2017. https://doi.org/10.1016/j.fuel.2017.07.046
[62] W. Tutak, K. Lukács, S. Szwaja, Á. Bereczky, "Alcohol-diesel fuel combustion in the compression ignition engine" Fuel, 154, 196–206, 2015. https://doi.org/10.1016/j.fuel.2015.03.071
[63] N. Yilmaz, F. M. Vigil, A. B. Donaldson, T. Darabseh, "Investigation of CI engine emissions in biodiesel-ethanol-diesel blends as a function of ethanol concentration" Fuel, 115, 790–793, 2014. https://doi.org/10.1016/j.fuel.2013.08.012
[64] S. Moon, T. Tsujimura, M. Oguma, Z. Chen, Z. Huang, T. Saitou, "Mixture condition, combustion and sooting characteristics of ethanol-diesel blends in diffusion flames under various injection and ambient conditions" Fuel, 113, 128–139, 2013. https://doi.org/10.1016/j.fuel.2013.05.060
[65] K. Baczewski, P. Szczawiński, M. Kamińska, "Experimental testing of influence of commercial depressants on diesel fuels low temperature properties" J. KONES, 22(1), 7–14, 2015. DOI: 10.5604/12314005.1161595
[66] J. Yanowitz, M. A. Ratcliff, R. L. McCormick, J. D. Taylor, M. J. Murphy, "Compendium of Experimental Cetane Numbers" (Technichal Report No. NREL/TP-5400-67585) National Renewable Energy Laboratory, Golden, CO, 2017.
[67] M. E. Tat, "Cetane number effect on the energetic and exergetic efficiency of a diesel engine fuelled with biodiesel" Fuel Process. Technol., 92(7), 1311–1321, 2011. https://doi.org/10.1016/j.fuproc.2011.02.006
[68] K. Takahashi, M. Sekimoto, H. Watanabe, T. Saito, R. Aoki, Y. Sakurai, T. Furuse, T. Sakuraba, T. Kezuka, Imai, S,. "Effects of cetane number and chemical components on diesel emissions and vehicle performance" Proceedings 20th World Petroleum Congress, Doha, Qatar, 2011. https://doi.org/10.4271/2009-01-2692
[69] J. P. Szybist, B. G. Bunting, "Cetane number and engine speed effects on diesel HCCI performance and emissions" SAE Technical Paper 2005-01-3723, 1334–1346, 2005. https://doi.org/10.4271/2005-01-3723
[70] J. Reijnders, M. Boot, P. de Goey, "Impact of aromaticity and cetane number on the soot-NOx trade-off in conventional and low temperature combustion" Fuel, 186, p. 24–34, 2016. https://doi.org/10.1016/j.fuel.2016.08.009
[71] E. Kurtz, C. J. Polonowski, "The influence of fuel cetane number on catalyst light-off operation in a modern diesel engine" SAE Int. J. Fuels Lubr. 10(3), 664–671, 2017. https://doi.org/10.4271/2017-01-9378.
[72] M. T. Chaichan, S. T. Ahmed, "Effect of fuel cetane number on multi-cylinders direct injection diesel engine performance and exhaust emissions" Al-Khwarizmi Eng. J., 8(1), 65–75, 2012.
[73] R. Cataluña, R. Da Silva, "Effect of cetane number on specific fuel consumption and particulate matter and unburned hydrocarbon emissions from diesel engines" J. Combust., 2012. http://dx.doi.org/10.1155/2012/738940
[74] O. C. Chukwuezie, N. R. Nwakuba, S. N. Asoegwu, K. N. Nwaigwe, "Cetane number effect on engine performance and gas emission: a review" American J. Eng. Res., 6(1), 56–67, 2017.
[75] S. M. Fakhr Hoseini, M. Dastanian, "Predicting pyrolysis products of PE, PP, and PET using NRTL activity coefficient model" J. Chem., 2013. http://dx.doi.org/10.1155/2013/487676
[76] API, HPV Testing Group Kerosene/jet fuel category assessment document. Submitted to the US EPA, by The American Petroleum Institute, Consortium Registration, 2010. 1100997.
[77] R. A. Harley, A. J. Kean, "Chemical Composition of Vehicle-Related Volatile Organic Compound Emissions in Central California Final Report (Contract 00-14CCOS) Prepared for: San Joaquin Valleywide Air Pollution Study Agency and California Air Resources Board, 2004.
[78] P. Ghosh, K. J. Hickey, S. B. Jaffe, "Development of a detailed gasoline composition-based octane model" Ind. Eng. Chem. Res., 45(1), 337–345, 2006. https://doi.org/10.1021/ie050811h
[79] P. Basu, Economic Issues of Biomass Energy Conversion, Springer, 2013 25–43. https://doi.org/10.1016/C2011-0-07564-6
[80] D. R. Nhuchhen, P. Basu, B. Acharya, "A comprehensive review on biomass torrefaction" Int. J. Renewable Energy Biofuels, 2014, 1–56, 2014. http://doi.org/10.5171/2014.506376
[81] M. Lapuerta, O. Armas, J. J. Hernández, "Potential for reducing emissions in a diesel engine by fuelling with conventional biodiesel and Fischer–Tropsch diesel" Fuel, 89(10), 3106–3113, 2010. https://doi.org/10.1016/j.fuel.2010.05.013
[82] G. Tüccar, E. Tosun, T. Özgür, K. Aydın, "Diesel engine emissions and performance from blends of citrus sinensis biodiesel and diesel fuel" Fuel, 132, 7–11, 2014. https://doi.org/10.1016/j.fuel.2014.04.065
[83] H. Raheman, S. V. Ghadge, "Performance of compression ignition engine with mahua (Madhuca indica) biodiesel" Fuel, 86(16), 2568–2573, 2007. https://doi.org/10.1016/j.fuel.2007.02.019
[84] O. Armas, K. Yehliu, A. L. Boehman, "Effect of alternative fuels on exhaust emissions during diesel engine operation with matched combustion phasing" Fuel, 89(2), 438–456, 2010. https://doi.org/10.1016/j.fuel.2009.09.022
[85] S. Godiganur, C. S. Murthy, R. P. Reddy, "Performance and emission characteristics of a Kirloskar HA394 diesel engine operated on fish oil methyl esters" Renewable Energy, 35(2), 355–359, 2010. https://doi.org/10.1016/j.renene.2009.07.007
[86] L. Zhu, W. Zhang, W. Liu, Z. Huang, "Experimental study on particulate and NOx emissions of a diesel engine fueled with ultra-low sulfur diesel, RME-diesel blends and PME-diesel blends" Sci. Total Environ., 408(5), 1050-1058, 2010. doi: 10.1016/j.scitotenv.2009.10.056.
[87] M.M. Rahman, M. Hj. Hassan, Md. A. Kalam, A. E. Atabani, L. A. Memon, S. M. Ashrafur Rahman, "Performance and emission analysis of Jatropha curcas and Moringa oleifera methyl ester fuel blends in a multi-cylinder diesel engine" J. Cleaner Prod., 65, 304–310, 2014. https://doi.org/10.1016/j.jclepro.2013.08.034
[88] J. B. Heywood, Internal Combustion Engine Fundamentals, McGraw-Hill, New York, 1988.
[89] P. Benjumea, J. R. Agudelo, A. F. Agudelo, "Effect of the degree of unsaturation of biodiesel fuels on engine performance, combustion characteristics, and emissions" Energy Fuels, 25(1), 77–85, 2010. https://doi.org/10.1021/ef101096x
[90] Ş. Altun, "Effect of the degree of unsaturation of biodiesel fuels on the exhaust emissions of a diesel power generator" Fuel, 117, 450–457, 2014. https://doi.org/10.1016/j.fuel.2013.09.028
[91] F. Lujaji, L. Kristóf, A. Bereczky, M. Mbarawa, "Experimental investigation of fuel properties, engine performance, combustion and emissions of blends containing croton oil, butanol, and diesel on a CI engine" Fuel, 90(2), 505–510, 2011. https://doi.org/10.1016/j.fuel.2010.10.004

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 97-103 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Toward a Smart Campus Using IoT: Framework for Safety and Security System on a University Campus

Alghamdi Abdullah[*,1], Mohammed Thanoon[2], Anwar Alsulami[3]

[1]*Information Systems Department, Najran University, Najran, Saudi Arabia*

[2]*Electrical and Computer Eng. Dept., Tennessee State University, Nashville, USA*

[3]*Department of Mechanical Eng., Jazan University, Jazan, Saudi Arabia*

ARTICLE INFO

Article history:
Received: 17 May, 2019
Accepted: 09 July, 2019
Online: 03 September, 2019

Keywords :
IoT
Smart Campus
Safety and Security

ABSTRACT

Recently, there is an enormous research on the smart campus concept due to the revolution of the IoT technologies. The motivation of this paper is to: reinforce the safety on campus, reduce the cost, and take one step forward toward a University smart campus. In this paper, we are not only proposing a framework that would act as an instantaneous responder, but we also provide a glimpse of the evolving research on smart campus. In addition, we explore the challenges, and highlight the future work regarding this on-the-spot responder system.

1. Introduction

The smart campus concept has been the main focus of many researchers recently due to the valuable insights gained toward developing smart cities. The university campus theoretically is a small city where it delivers variant services to variant users. There are several factors that attract the investigators to study the smart campus including: delivering high quality services, protecting the environment, and saving the cost. The internet of things is a fundamental part of the smart campus, and it is inescapable getting it invoked.

The internet of things is a communication paradigm that gained its tenacity from its capability of connecting variety of everyday life objects to the internet. These objects include, but not limited to, sensors, robots, security locks, alarms, drones, appliances, smart grid systems, office equipment and so on. Even though IoT is in its early stages, there are many applications and standardization that has been developed in many domains including: home automation, smart grids, water and waste management, traffic control, smart vehicles, healthcare assistance, and industrial automation. Moreover, the realization of the IoT network is facing two main challenges. First, technical challenges because of IoT novelty and heterogeneous nature. Second, business challenges due to the lack of complete and approved business model that would encourage investments.

In this paper, we are proposing a safety and security framework based on the IoT which would enhance the safety on a university campus. The work in this paper is not only an extension of the investigation originally presented at the 2016 International Conference on Future Internet of Things and Cloud (FiCloud) [1], but we also propose a new system, and identify the limitations and research opportunity in this field.

The safety on a university campus is a growing concern among the campus community across the U.S. and across so many other countries due to the horrific crimes and mass shooting committed, lately. Regardless of whether the campus is rural or urban, large or small, significant number of people on campus seems to be reluctant to walk out at night alone [2]. According to Clery Center for security on campus [3], the crimes committed by students on campus is about 80% of the overall crime number. For example, Tennessee State University (TSU) is considered about an average in crime rates [4]; However, there were recent cases which saddened TSU community such as the shooting in 2015 that killed one, and injured three students [5]. Figure 1 shows statistics of the crimes committed on post high school education campuses in the United States over the period of 2015 to 2017 [6]. Our proposed system will not only enhance the safety, but also it would reduce the cost, make efficient use of resources, and

*Alghamdi Abdullah, Assistant Professor, IS Dept., NU, Najran, Saudi Arabia.
Email: aaalghamdi@nu.edu.sa

become one step closer toward a smart campus by utilizing the recent advances in technology.

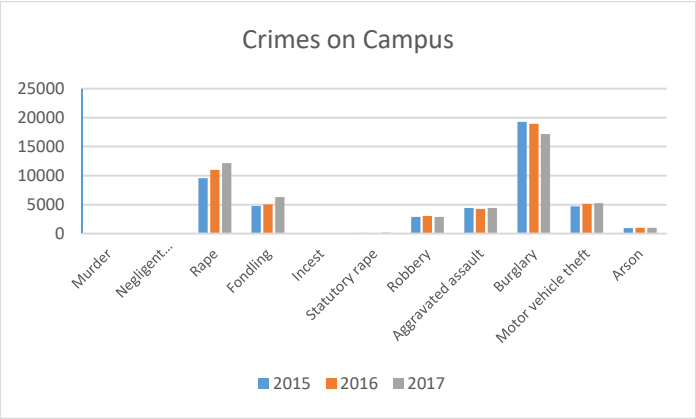


Figure 1 Statistics of The Crimes Committed on more than 11000 Campuses [6].

2. Smart Campus and Its Applications

The smart campus market springs from being a diverse environment regarding the services conveyed and the recipients of these services. The impact of conveyed services is not limited to the academic aspect, but also the social, financial, and environmental aspects. We can categorize the current research in smart campus into four main areas: intelligent buildings, campus smart grid, learning environment, and other applications. Figure 2 shows the context of a smart campus, its impacts, and many examples of its applications. On the rest of this section, we will present the research being done toward achieving a smart campus.

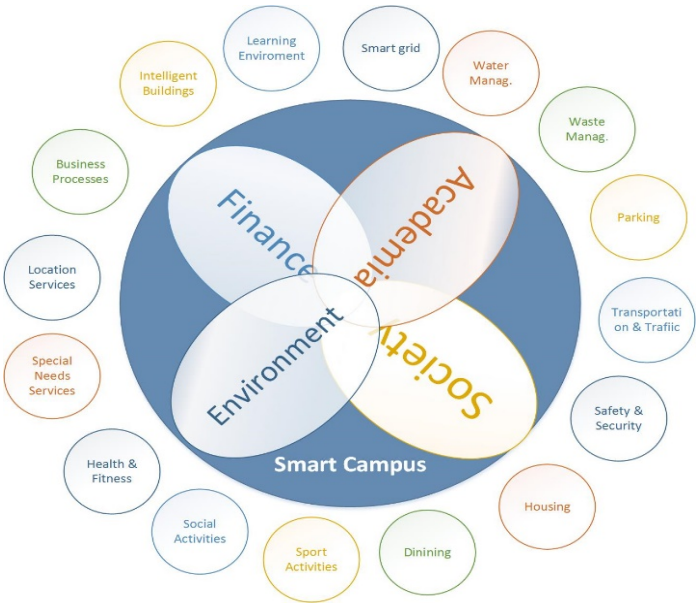


Figure 2 Impacts and Applications of Smart Campus Concept [1].

2.1. The Intelligent Buildings

Given that there is no de facto guide which would lead the people in the construction industry on how to build an intelligent building, a novel approach would be the plan to achieve the stakeholders' goals [7]. The planners need to focus on two main aspects. First, Information and Communication Technology (ICT) infrastructure where at least one fast backend network is needed for both business and building systems. Servers and variant sensors are essential part of the automation process of buildings. Sensors routine includes, detecting carbone dioxide levels in buildings, monitoring building stress, adjusting temperature, turning off the services in empty spaces, and measuring humidity and pollution levels [8,9]. Second, reporting and collaboration between various department is crucial to improve business processes and workflow, and it can be done by sharing information and reports, and enhancing communication [7]. The reports include building occupancy, staff attendance, patterns of utility usage, real-time warnings, energy usage, and so on. Knowing that the productivity gets increased by boosting the comfort of the community on campus, the stakeholders would make that as one of their goals [10].

2.2. Campus Smart Grid

Smart grid "is an opportunity to use new information and communication technologies (ICTs) to revolutionize the electrical power system" [11]. There are many universities adopting smart grids such as Oregon State University, which integrated Synchrophasor technology to have better monitoring and alarming system; however, the optimized results are not realistic due to the fact that this technology is relatively new and needs further studies [12]. In case of failures with SCADA, Santos *et al.* [13] suggests an automation process of power restoration using Open Platform Communications (OPC) protocol, and to characterize detection, isolation, and self-healing. Examining demand patterns, implementing a self-organized mapping (SOM), which is favored load profiling algorithm, and analyzing the current grid to provide a methodology and recommendations for smart microgrid in a university campus are discussed in [14,15]. Web-based systems on a cloud platform are introduced in: [16] for the analysis of energy consumption and behavior patterns, [17] for the purpose of controlling the demand response, and [18] for forecasting of future demand using machine learning (ML) models. Bracco *et al.* [19] propose a quantifying method for the usage of primary energy, predict CO2 emissions cut, and assess the cost reduction at the University of Genoa campus. Finally, IoT made it possible to control and monitor the energy consumption by users, identify energy draining devices, and suggests actions to optimize their behavior.

2.3. The Learning Environment

Given that the primary ambition of e-learning systems is to provide an articulated learning environment for learners based on their aspirations, knowledge, and talent, Wang *et al.* [20] characterize the learning process in an intelligent environment as a recursive process of four stages: learning, assessment, interaction, and analysis. Several papers published in this area including [21] which proposes students profiling based on their behavior over on-campus social network to provide a context-based personalized learning experience to achieve a ubiquitous learning (uLearning); however, additional study in terms of

learners' behavior and content design would exaggerate the benefit of such system. Microsoft and MIT started a research collaboration to achieve an intelligent campus, so called (MIT iCampus) [22]. This collaboration resulted in developing several intelligent systems including: class communicator system (CCS) to overcome communication issues between instructors and students, and class learning partner (CLP), and to provide an exercises and instant feedback during the class. Finally, given that mobile phones, tablets, laptops, and other wearable devices are among the most prominent devices that have influenced this universe [23], and utilizing them in education is inevitable.

2.4. Other Applications

Even though they do not get enough attention, there are plenty of other applications where IoT can be utilized to dramatically change campus to a smart one such as: waste and water management, parking, voting, safety and security, and so on. Regarding waste management, many papers published recently including [24-26] which generally suggest planting sensors at the bins, and waste trucks that collect real-time data for analysis, and a system would use the collected data to suggest better cleaning schedule and a superior and cost-effective route for waste contractors.

One the other hand, Mudumbe and Abu-Mahfouz [27] present a user-centric smart water management system which represent consumption in visual graphs to increase awareness. Gabrielli *et al*. [28] describe a sustainable prototype for smart metering devices and the associated network. Despite the feasibility of the project, the focus of their experiment is to build a sustainable metering device. Moreover, the authors [29,30] demonstrate: a parking guidance system for the parking building, and an electronic voting system that uses RFID technology for the authentication process, respectively.

Keeping in mind that most of recent work is focused on energy, learning environment, and intelligent buildings, the safety and security did not attract the attention of researchers; however, the intrinsic need of: better safety, instantaneous response to emergency, and cost effective solution on campuses actuated our team to investigate this area.

3. Safety and Security System

In this section, we will describe our system in details. At first, we will show the overall architecture of the system followed by the technique that will be used to deploy the sensors. Considering four different types of undesirable incidents: gun possession, assault, burglary and firing arms, we will show how to detect such suspicious activity. After that, we will show how to detect and track the suspect until the backup team arrives to the scene.

3.1. System Architecture

The safety and security system work with a variety of devices including: cameras, microphone sensors, glass break sensors, Raspberry Pi boards (RPis), a drone, and a server connected to the safety and security Operation Control Center (OCC) on campus. Figure 3 shows the general architecture of the system. To avoid overwhelming one server by making it process all the data collected by sensors across the campus, we would use the RPis to accomplish a distributed data processing architecture. Taking advantage of its cheap cost, the RPis would be used to process images, voices and other collected data by attached sensors. In case of detecting suspicious activity, the RPi would notify the server which will identify the location of the RPi using table for mapping the RPi's ID and its predefined location. The server would trigger the alert, identify the fastest route to the incident location, and instruct the drone to observe the situation, and wait for further actions.

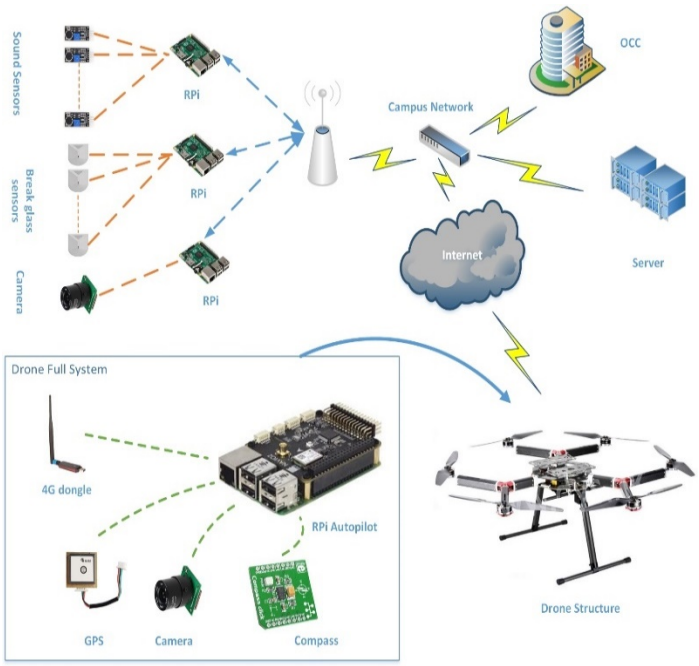


Figure 3 System Architecture

A sophisticated and robust drone need to be used as an instantaneous responder to observe the situation, and send live updates to the OCC for assessment and evaluation. The drone's autopilot would be attached to: a 4G dongle for internet connection, a GPS, a camera, and a compass.

3.2. Sensors Deployment

Since we have different types of sensors and our system depend on them, we need to deploy them carefully. The deployment technique is planned attentively to achieve two goals: to be cost effective regarding the number of sensors, and to avoid degrading the effectiveness of the system by deploying less sensors than required. The cameras would be deployed at parking lots, sidewalks, and buildings' entrances. The glass break sensors need to be deployed inside the buildings. Because one sensor covers around 25 feet, we would deploy one sensor per room wherever applicable. Finally, we decided to choose a microphone that covers 20 to 30 feet in a spherical area such as QSPMIC microphone by Q- See.

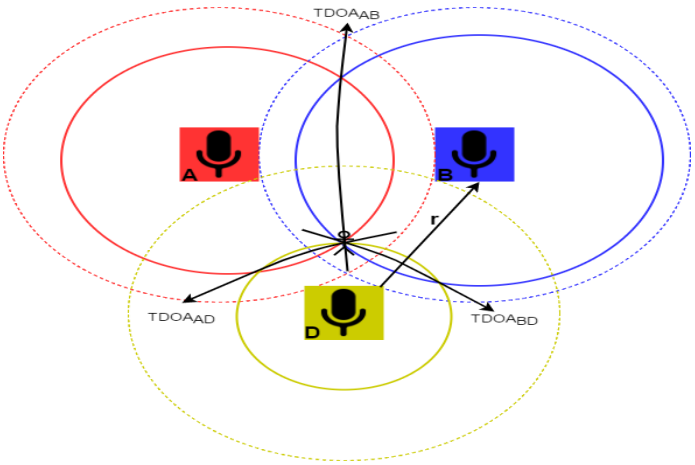


Figure 4 Sound Sensors Deployment

Figure 4 shows the way of planting the sound sensors where every sensor is planted at the end of the previous sensor coverage. Thus, the horizontally and vertically distance between sensors is equal to the radius of the coverage area (r). The sound source location can be determined using the Time Difference of Arrival (TDOA) or Multilateration (MLAT) method [31]. TDOA is an algorithm used to determine the location of the sound source by calculating the differences in arrival time of the sound at spatially separated sensors. The time difference of arrival of sound at different sensors must be known and processed at the central location. When the possible distance of the source from each sensor is drawn, it creates a hyperbolic curve. The location of the sound source lies in the intersection of all the hyperbolas from different sensors. This method works quite accurately for relatively static sources. If dynamic sources to be tagged, modified TDOA can be used [32].

3.3. Suspicious Activity Detection

The proposed system considers four different undesirable incidents to prove its effectiveness which are: gun possession, assault, burglary, and firing arms. The following paragraphs will study these scenarios in details.

Gun possession would be caught using the cameras distributed all over the campus parking lots, entrances, and sidewalks. The Raspberry Pies (RPis) attached to cameras would be able to detect guns automatically using Convolutional Neural Networks (CNNs) [33]. CNN is the state of the art technology in computer vision and it has been proven as a reliable architecture in object detection, giving its high accuracy with sufficient training. CNN is a special case of Neural Networks that apply the convolution operation to an input image for the purpose of feature extraction; however, the usage of the convolution operation demands high computational time, which is a challenge in our environment, knowing that we are using limited, low-energy devices. Fortunately, Iandola *et al.* [34] were able to achieve a smaller CNN architecture that performs the same accuracy as the AlexNet-level on ImageNet with fifty times fewer parameters, and they named it SqueezeNet. As a result, the SqueezeNet can be used on almost any controller or embedded computers. On RPis, it takes SqueezeNet approximately 2 seconds to process an image, which is enough time to: detect a person with a gun on a sidewalk, issue an alarm on the building, automatically close the inner doors of buildings, and notify the safety and security OCC on campus. Using a significant number of images of different types of guns, the SqueezeNet would be trained. The trained model would be deployed on the RPis wherever the cameras are applicable. After the detection of the gun, the RPis would alert the server. Using the predefined mapping between RPis and locations, the server would be able to instantaneously deduce the location of the gunman, plan the fastest route for drone, and instruct the drone to find and tracking the gunman until a backup team arrives to the scene.

There are some cases where the gun would not be detected by the distributed cameras. Therefore, we introduce another method for detection which is by detecting the sounds of firing arms. Usually, firing the arms and assault crimes are accompanied by sounds such as the sounds of gunshots and yelling for help. In case of any of these undesirable incidents happened, the microphones attached to RPis would catch the accompanied signals, and pass it to the neural network to identify if there is any suspicious activity. Since the CNN is proven to be effective in speech recognition [35], likewise, the SqueezeNet architecture would be used here too. Gunshots and yelling for help spectrograms would be used to train the network, and the pre-trained model would be deployed on the RPis. Similar to the previous scenario when the detection happens, the RPis would notify the server. Then, the server would identify the location of the incident and instruct the drone to observe the situation.

For burglary detections, we are using a glass break sensors which are connected to RPis. By listening to acoustical frequencies generated when a glass gets broken, the sensor would alert the RPi over the wireless which would lead to issuing an alarm to the server. Because the RPi mapped the sensor with its location on the building, the server would send the drone to that location to give real-time updates of the broken window. This would be important in case the burglar tried to run away because the drone can always detect and track people on the scene. Meanwhile, the OCC room should have sent a backup team to that room to investigate the situation.

3.4. Target Detection and Tracking

As soon as the drone arrives into the scene, it has to detect and track the suspect. To add more capability, our algorithm would allow two methods of detection: manually and automatically based on the situation. Figure 5 shows the general flowchart of target detection and tracking process. Upon the arrival of drone, it would start: capturing videos frames, notifying the OCC with updates, and analyzing the captured frames for detection of the suspect. Knowing that CNN requires high computation demand, Histogram Oriented Gradient (HOG) [36] would be used to detect upright persons in the scene. Because we need to reduce the computation overhead of the RPi on drone, we would analyze only five frames out of the twenty-five frames per

second that the camera can detect. If there is more than one upright person on the scene, the system would allow the OCC to decide which one to track. Otherwise, the drone will track the detected target. Regardless of the method used to detect the target, the drone will keep the target on its camera frame center, while the OCC will get real-time updates of the location which would be shown on an interactive map. Because we don't want to overwhelm the drone's connection to the internet, the OCC will get few captured frames only.

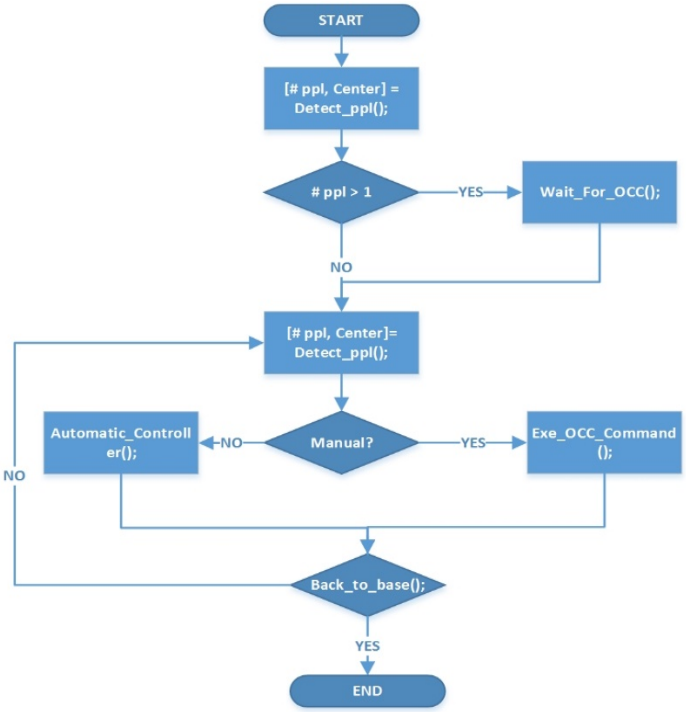


Figure 5 Target Detection and Tracking Flowchart

For the tracking, our system would allow the manual tracking of the suspect where the OCC officer can remotely control the drone without the need for frames analysis; however, this would require a high speed internet because the OCC needs a live streaming to be able to remotely control the drone. In case of automatic tracking, the optical flow method and its estimation using the Lucas-Kanade method will be used to track the suspect at the first frame only of every second [37,38]. The method works as the following: consider a pixel *I(x,y,t)* at the first frame, and it moves by distance *dx, dy,* in the frame that is compared to which is taken at *dt*. Because we are tracking the same person on real time, we assume the intensity does not change.

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \quad (1)$$

by taking Tylor series approximation, and dividing by *dt*, we get:

$$f_x u + f_y v + f_t = 0 \quad (2)$$

$$f_x = \frac{\partial f}{\partial x}; \quad f_y = \frac{\partial f}{\partial y} \quad (3)$$

$$u = \frac{dx}{dt}; \; v = \frac{dy}{dt} \quad (4)$$

Then, using the Lucas-Kanade method, the final solution will be:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum_i fx_i^2 & \sum_i fx_i fy_i \\ \sum_i fx_i fy_i & \sum_i fy_i^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i fx_i ft_i \\ -\sum_i fy_i ft_i \end{bmatrix} \quad (5)$$

Then, the drone will move to the desired direction to make u, and v in its frame center. This method will be repeated once every second. Because we are capturing from the drone, the capture has a wide area where the suspect is in the center of the frame. It is less likely that the suspect will be able to go beyond the area of the drone's camera in one second. So, we are processing one frame per second only, and to avoid overwhelming the RPi with computation demands. Figure 6 illustrates the detection and tracking process, and the movement of the drone.

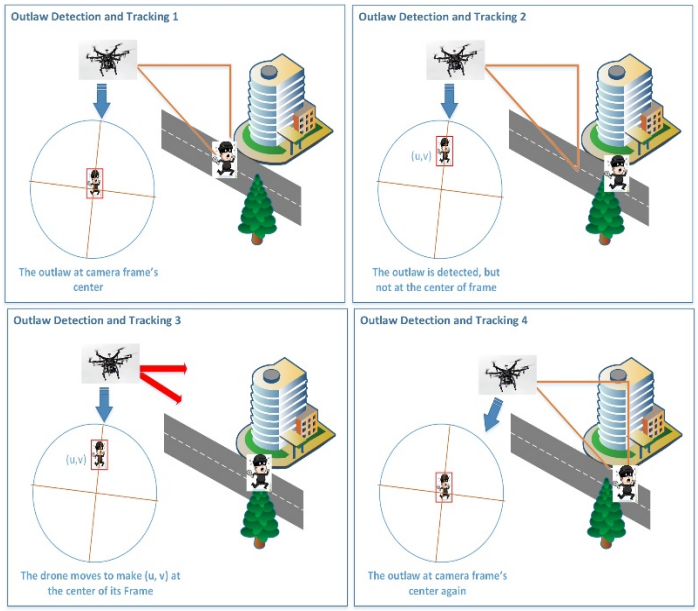


Figure 6 Target Detection and Tracking

4. Challenges and Future Work

Basically, the proposed safety and security framework inherited the obstacles that are facing the smart campus market which we can summarize in three areas: technical, financial, and political.

The technical barriers can be observed from the following perspectives: security, privacy, and configuration [1]. Being able to planet a massive number of low-energy and wirelessly connected devices on campus and meeting the security requirements is a problematic task due to the heterogeneous and intensive communication environment. Giving that we are utilizing various types of sensors, the privacy of people on campus is an issue where we need to be absolute that nobody can misuse the system to invade the people's privacy. Adopting the safety and security system on campus would result in using hundreds if not thousands of sensors which can be an enormous burden to configure them manually. Certainly, new ways of the automated configuration of IoT devices need to be thoroughly investigated.

In general, the smart campus market is facing financial difficulties giving that the limited resources of universities, regardless of immature experiences here and there. Even though the proposed system would save the university tons of many regarding the number of security officers on the long term, the initial investment is an impediment to the implementation of this system.

Regarding political hindrances, this system would need a waiver of some restrictions for unmanned aerial vehicles (UAV) by the Federal Aviation Administration (FAA), in the United States or such agencies in other countries, such as: the maximum altitude of 400 feet above ground level, visual line of sight, daylight only operations and some others [39]. Also, the system may face forceful opposition from people on campus and nearby neighborhoods regarding the invasion of their privacy. Even though it is almost inconceivable, the opposition of anti-tech security officers is a possible hindrance that needs to be considered.

After finalizing the implementation, our work opens the doors for researchers to improve the capability of this system to achieve optimum performance. The detection of outlaws during night-time and tracking them is a possible area of improvement. Moreover, the profiling of crimes respecting the hotspots, timings, and frequent suspicious activities using data collected by sensors would provide copious information about the crimes. The refined information would lead to crimes' prediction and prevention, eventually. Reducing the time between incident and arrival of drone is another area of improvement. Arming the drone with stun gun shocks could be considered for improvement. In addition, finding ways to avoid outlaws targeting the drone such as working on a bulletproof drone structure and trying to balance that with the speed of the drone is another area of improvement.

5. Conclusion

After presenting the recent work in smart campus technologies using IoT, we proposed a safety and security framework tailored to fit any university campus that would like to take advantage of recent advances in technology. The proposed framework would act as an instantaneous responder to incidents that could happen on campus. The proposed system is a good tool that would not only detect the incidents, but also it can track the outlaws even outside the campus until a backup team arrive to the scene. Also, we discussed the challenges of the system from three different perspectives: technical, and financial, and political. Finally, we highlighted the future work for this system to achieve its optimum performance.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] A. Alghamdi and S. Shetty, "Survey Toward a Smart Campus Using the Internet of Things," in *2016 IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud)*, 2016, pp. 235–239.

[2] A. L. Ball, "Students Fear Venturing Out Alone at Night on Campus.," *The New York Times*, 20-Jul-2012.

[3] "Clery Center For Security On Campus." [Online]. Available: http://clerycenter.org/. [Accessed: 25-Dec-2018].

[4] College Factual, "Tennessee State University Crime," *College Factual*, 20-Feb-2013. [Online]. Available: http://www.collegefactual.com/colleges/tennessee-state-university/student-life/crime/. [Accessed: 25-Dec-2018].

[5] N. N. Alund, "Two men in custody in 2015 TSU fatal campus shooting," *The Tennessean*. [Online]. Available: http://www.tennessean.com/story/news/crime/2016/08/23/two-people-custody-2015-tsu-campus-shooting/89211076/. [Accessed: 25-Dec-2018].

[6] U.S. Department of Education, "Campus Safety and Security." [Online]. Available: http://ope.ed.gov/campussafety/#/. [Accessed: 25-Dec-2018].

[7] S. Hipwell, "Developing smart campuses #x2014; A working model," in *2014 International Conference on Intelligent Green Building and Smart Grid (IGBSG)*, 2014, pp. 1–6.

[8] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for Smart Cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, Feb. 2014.

[9] J. P. Lynch, "A Summary Review of Wireless Sensors and Sensor Networks for Structural Health Monitoring," *The Shock and Vibration Digest*, vol. 38, no. 2, pp. 91–128, Mar. 2006.

[10] W. Kastner, G. Neugschwandtner, S. Soucek, and H. M. Newmann, "Communication Systems for Building Automation and Control," *Proceedings of the IEEE*, vol. 93, no. 6, pp. 1178–1203, Jun. 2005.

[11] A. B. M. S. Ali, Ed., *Smart Grids Opportunities, Developments, and Trends*. London: Springer London, 2013.

[12] C. Huo, J. Song, K. Wagner, G. Harold, and E. Cotilla-Sanchez, "Integrating synchrophasor technology with the Oregon State University campus smart grid project," in *2014 IEEE Conference on Technologies for Sustainability (SusTech)*, 2014, pp. 125–129.

[13] L. A. Santos *et al.*, "Towards a smart grid to the university campus of the federal university of Cear??," in *2015 IEEE 13th Brazilian Power Electronics Conference and 1st Southern Power Electronics Conference (COBEP/SPEC)*, 2015, pp. 1–6.

[14] I. P. Panapakidis, T. A. Papadopoulos, G. C. Christoforidis, and G. K. Papagiannis, "Analysis of the electricity demand patterns of a building in a university Campus," in *2013 12th International Conference on Environment and Electrical Engineering (EEEIC)*, 2013, pp. 382–387.

[15] R. M. González, T. A. J. van Goch, M. F. Aslam, A. Blanch, and P. F. Ribeiro, "Microgrid design considerations for a smart-energy university campus," in *Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), 2014 IEEE PES*, 2014, pp. 1–6.

[16] V. Nikolopoulos, G. Mpardis, I. Giannoukos, I. Lykourentzou, and V. Loumos, "Web-based decision-support system methodology for smart provision of adaptive digital energy services over cloud technologies," *IET Software*, vol. 5, no. 5, pp. 454–465, Oct. 2011.

[17] H. Kim, Y. J. Kim, K. Yang, and M. Thottan, "Cloud-based demand response for smart grid: Architecture and distributed algorithms," in *2011 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2011, pp. 398–403.

[18] Y. Simmhan *et al.*, "Cloud-Based Software Platform for Big Data Analytics in Smart Grids," *Computing in Science Engineering*, vol. 15, no. 4, pp. 38–47, Jul. 2013.

[19] S. Bracco, F. Delfino, F. Pampararo, M. Robba, and M. Rossi, "Economic and environmental performances quantification of the university of Genoa Smart Polygeneration Microgrid," in *Energy Conference and Exhibition (ENERGYCON), 2012 IEEE International*, 2012, pp. 593–598.

[20] M. Wang and J. W. P. Ng, "Intelligent Mobile Cloud Education: Smart Anytime-Anywhere Learning for the Next Generation Campus Environment," in *2012 8th International Conference on Intelligent Environments (IE)*, 2012, pp. 149–156.

[21] Y. Atif and S. Mathew, "A Social Web of Things Approach to a Smart Campus Model," in *Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing*, 2013, pp. 349–354.

[22] "iCampus," *iCampus*. [Online]. Available: http://icampus.mit.edu/. [Accessed: 05-Oct-2018].

[23] V. G. Cerf, "Prospects for the Internet of Things," *XRDS*, vol. 22, no. 2, pp. 28–31, Dec. 2015.

[24] T. Anagnostopoulos, A. Zaslavsky, and A. Medvedev, "Robust waste collection exploiting cost efficiency of IoT potentiality in Smart Cities," in

2015 International Conference on Recent Advances in Internet of Things (RIoT), 2015, pp. 1–6.

[25] F. Folianto, Y. S. Low, and W. L. Yeow, "Smartbin: Smart waste management system," in *2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, 2015, pp. 1–2.

[26] T. Anagnostopoulos, A. Zaslavsy, A. Medvedev, and S. Khoruzhnicov, "Top – k Query Based Dynamic Scheduling for IoT-enabled Smart City Waste Collection," in *2015 16th IEEE International Conference on Mobile Data Management (MDM)*, 2015, vol. 2, pp. 50–55.

[27] M. J. Mudumbe and A. M. Abu-Mahfouz, "Smart water meter system for user-centric consumption measurement," in *2015 IEEE 13th International Conference on Industrial Informatics (INDIN)*, 2015, pp. 993–998.

[28] L. Gabrielli, M. Pizzichini, S. Spinsante, S. Squartini, and R. Gavazzi, "Smart water grids for smart cities: A sustainable prototype demonstrator," in *2014 European Conference on Networks and Communications (EuCNC)*, 2014, pp. 1–5.

[29] W. Atmadja, J. Yosafat, R. A. Setiawan, and I. I. Irendy, "Parking guidance system based on real time operating system," in *2014 International Conference on Industrial Automation, Information and Communications Technology (IAICT)*, 2014, pp. 5–8.

[30] A. Saad, M. I. M. Roseli, and M. S. Zullkeply, "A Smart e-Voting System Using RFID Authentication Method for a Campus Electoral," in *Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication*, New York, NY, USA, 2014, pp. 31:1–31:7.

[31] F. Gustafsson and F. Gunnarsson, "Positioning using time-difference of arrival measurements," in 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03), 2003, vol. 6, p. VI-553-6 vol.6.

[32] K. C. Ho and W. Xu, "An accurate algebraic solution for moving source location using TDOA and FDOA measurements," IEEE Transactions on Signal Processing, vol. 52, no. 9, pp. 2453–2463, Sep. 2004.

[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," presented at the Advances in Neural Information Processing Systems.

[34] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," *arXiv:1602.07360 [cs]*, Feb. 2016.

[35] O. Abdel-Hamid, A.-R. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolutional Neural Networks for Speech Recognition," *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 22, no. 10, pp. 1533–1545, Oct. 2014.

[36] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01, Washington, DC, USA, 2005, pp. 886–893.

[37] S. Baker and I. Matthews, "Lucas-Kanade 20 Years On: A Unifying Framework," *International Journal of Computer Vision*, vol. 56, no. 3, pp. 221–255, Feb. 2004.

[38] OpenCV, "OpenCV: Optical Flow," *Open Source Computer Vision*. [Online]. Available: http://docs.opencv.org/trunk/d7/d8b/tutorial_py_lucas_kanade.html. [Accessed: 25-Dec-2016].

[39] Federal Aviation Administration, "Summary of Small Unmanned Aircraft Rule (Part 107)," FAA, Washington, DC 20591, Jun. 2016.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 104-109 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

pH Neutralization Process Control based on Active Disturbance Rejection Control

Mohanad Hamad Eljack Elameen*, Mona Bahaeldin Musa Hago, Wen Tan

School of Control and Computer engineering, North China Electric Power University, 102206, China

ARTICLE INFO

Article history:
Received: 06 June, 2019
Accepted: 16 August, 2019
Online: 03 September, 2019

Keywords:
ADRC
Extended state observer (ESO)
rADRC
Nonlinear process
pH neutralization process

ABSTRACT

This paper considers active disturbance rejection control (ADRC) as a new control approach for pH neutralization process. The pH process is challenging to control due to its sensitivity to disturbance, nonlinear dynamics, and time-varying characteristics. Special attention is paid to the impact of disturbances typically exists in the chemical process. First, the pH neutralization process nonlinear model is presented, then the active disturbance rejection control is described briefly and a pH neutralization process closed-loop control system is designed on its bases, and investigated with simulation. More robust and better disturbance rejection performances are achieved compared to the control technique of feedback linearization (FL).

1. Introduction

This paper is an extension of work presented initially in 2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP) [1]. Control of pH neutralization in the chemical and biotechnological industries is a typical process [2]. The pH of this effluent must be retained within environmental boundaries, before discharging the effluent streams from wastewater treatment plants. Tight pH control is also critically required in pharmaceutical production. However, owing to the high nonlinearity of the pH process, robust and high performance pH process control is often difficult to obtain [3-5]. The process gain of these processes can differ from a small range of pH values to several orders of magnitude. Thus it can exhibit severe static nonlinear behavior.

So far, many authors have proposed, developed, and implemented different control techniques to overcome this problem. The classical PID controller is the first technique applied to control the pH [6-8]. In the field of industrial applications, the PID controller is considered as the most popular controller types, but it has been unable to satisfy the growing demand for efficiency, speed, and accuracy. Nonlinear control techniques also have been applied [9-21]. Other works involve intelligent control technique [22, 23]. In general, most of the above methods require an adequate model, which could not be easy to obtain, add to that the complexity of the controller design based on the above approaches. Hence, pH neutralization process requires a new approach to control it.

In [24-26], Han introduced the powerful control technique active disturbance rejection control (ADRC) and simplified and identified in [27]. The states and disturbance can be estimated in the ADRC before an error can occur, which can ultimately attain early compensation and disturbance resistance. Since it was introduced, the ADRC method has gained worldwide attention and broad implementation and has become common after its linear form has appeared [28]. In many fields, ADRC has been implemented effectively, benefiting from its full control characteristics as fast system response, large scale adaptation, and small overshoot [29].

In this technique, a unique state observer is used to estimate what is called total disturbance. The uncertainties in the dynamics of the internal process and the disturbances of the output are considered as total disturbances. This observer is called the extended state observer (ESO), and it can track the state-space model's extended state, which can achieve this goal. The plant's state-space model should include only the model estimate. A linear model in a state-space form may be an estimate for a nonlinear system [29].

This work proposes a different control technique based on estimating the disturbances and actively rejecting it for pH neutralization process. All the disturbances typically presented in pH process are lumped together in one signal called total disturbances. The extended state observer tracks these disturbances and the control law cancels it.

The organization of this paper is as follows. The pH process's model is given in section 2. Section 3 briefly presents the technique

* Mohanad Hamad Eljack Elameen, mohanadeljack@gmail.com

of feedback linearization. Section 4 shows the ADRC control system for the pH process. Section 5 gives the simulation results, and section 6 gives a general conclusion.

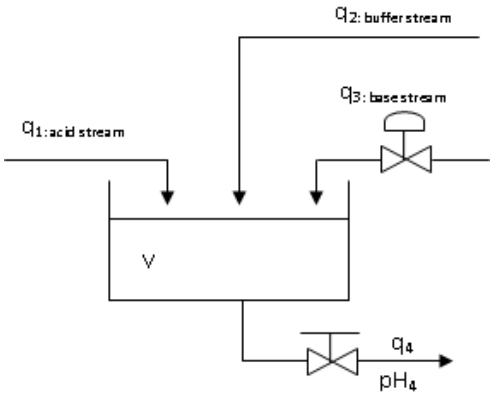
2. Model of pH Process

Figure 1 shows a schematic diagram describing the pH process. The process consists of an acid flow rate (q_1), buffer flow rate (q_2), and the base flow rate (q_3) mixed in the tank. pH_4 represents the output of the process, where q_3 represents the control input. By using the conservation equations of the reaction invariants and equilibrium relations, a dynamic model has been derived. Set tank volume (V), Ideal mixing, constant density, and full ion solubility are considered as assumptions for modeling. The system's chemical reactions are:

$$H_2CO_3 \leftrightarrow HCO_3^- + H^+,$$

$$HCO_3^- \leftrightarrow CO_3^{2-} + H^+,$$

$$H_2O \leftrightarrow OH^- + H^+$$



The reaction's equilibrium constants are

$$K_{a1} = \frac{[HCO_3^-][H^+]}{[H_2CO_3]},$$

$$K_{a2} = \frac{[CO_3^{2-}][H^+]}{[HCO_3^-]},$$

$$K_w = [H^+][OH^-]$$

The chemical balance is formed using the reaction invariant concept [30]. The invariants concentrations are (i=1,2,3,4):

$$W_{ai} = [HCO_3^-]_i + [H^+]_i - 2[CO_3^{2-}]_i - [OH^-]_i,$$

$$W_{bi} = [HCO_3^-]_i + [CO_3^{2-}]_i + [H_2CO_3]_i,$$

where W_{ai} is a charge-related quantity, and W_{bi} is the carbonate ion concentration, and they are independent of the extent of the reaction. The concentration of the hydrogen ion is calculated from:

$$W_{ai} + W_{bi}\frac{\frac{K_{a1}}{[H^+]_i} + \frac{2K_{a1}K_{a2}}{[H^+]_i^2}}{1 + \frac{K_{a1}}{[H^+]_i} + \frac{K_{a1}K_{a2}}{[H^+]_i^2}} + \frac{K_w}{[H^+]_i} - [H^+]_i = 0.$$

The hydrogen ion $[H^+]_4$ with the following relationship calculates the pH of the effluent stream:

$$pH_4 = -log_{10}([H^+]_4)$$

The dynamic model is given by the vessel's material balance:

$$V\frac{dW_{a4}}{dt} = q_1(W_{a1} - W_{a4}) + q_2(W_{a2} - W_{a4}) + q_3(W_{a3} - W_{a4}),$$

$$V\frac{dW_{b4}}{dt} = q_1(W_{b1} - W_{b4}) + q_2(W_{b2} - W_{b4}) + q_3(W_{b3} - W_{b4}).$$

This put the process model in a state-space form as

$$\dot{X} = f(x) + g(x)u + p(x)d \quad (1)$$

where

$$u = q_3, \quad x = [W_{a4}\ W_{b4}]^T, \quad d = [W_{a1}\ W_{b1}]^T,$$

$$f(x) = \begin{bmatrix} -\frac{q_1}{V}x_1 + \frac{q_2}{V}(W_{a2} - x1) \\ -\frac{q_1}{V}x_2 + \frac{q_2}{V}(W_{b2} - x2) \end{bmatrix},$$

$$g(x) = \begin{bmatrix} \frac{1}{V}(W_{a3} - x1) \\ \frac{1}{V}(W_{b3} - x2) \end{bmatrix}, \qquad p = \begin{bmatrix} \frac{q_1}{V} & 0 \\ 0 & \frac{q_1}{V} \end{bmatrix}.$$

Note that Eq. (1) is linear in p, and the unknown parameters are presented in the vector d. The relation between the state variables x and the output y is nonlinear as shown by

$$c(x, y) = 0, \quad (2)$$

where

$$c(x, y) = x_1 + 10^{y-14} - 10^{-y} + x_2 c_{x2}, \quad (3)$$

$$c_{x2} = \frac{1 + 2 \times 10^{y-pk_2}}{1 + 10^{pk_1 - y} + 10^{y-pk_2}},$$

$$pk_1 = -log_{10}k_{a1}, \quad pk_2 = -log_{10}k_{a2}.$$

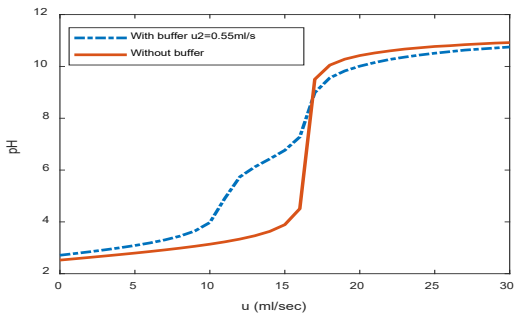


Figure 2: The dynamics of the pH process

Table 1 shows the nominal values of the parameters.

Table 1. Model Parameters Nominal Values

Parameter	Nominal value
V	2900 ml
K_{a1}	4.47×10^{-7}
K_{a2}	5.62×10^{-11}
$[q_1]$	0.003 M HNO3
	5×10^{-5} M H2CO3
$[q_2]$	0.03 M NaHCO3
$[q_3]$	0.003 M NaOH
	5×10^{-5} M NaHCO3
q_1	16.6 ml/s
q_2	0.55 ml/s
q_3	15.8 ml/s
pH_4	7.00
W_{a1}	0.003 M
W_{b1}	5×10^{-5} M
W_{a2}	-0.03 M
W_{b2}	0.03 M
W_{a3}	-3.05×10^{-3} M
W_{b3}	5×10^{-5} M
W_{a4}	-4.50×10^{-4} M
W_{b4}	5.5×10^{-4} M

Figure 2 shows the static dynamics of the pH process, which shows strong nonlinearity.

3. Feedback Linearization Technique

For the severely nonlinear system expressed in (1-3), a nonlinear controller based on feedback linearization (FL) technique is first designed.

Consider the following nonlinear system

$$\dot{x} = f(x) + g(x)u, \quad y = h(x) \tag{4}$$

The objective is to find a linear differential relation between a new input v and the output y. The system has a relative degree r if

$$L_g L_f^i = 0 \qquad 0 \le i \le r - 2 \tag{5}$$

$$L_g L_f^{r-1} h(x) \neq 0$$

where Lf in the direction of f is the Lie derivative. Suppose the system above has a relative degree of r and is linearizable, then the input transformation

$$u = \frac{v - L_f^r h}{L_g L_f^{r-1} h} \tag{6}$$

hence, the relation between y and v is linear as

$$y^{(r)} = v \tag{7}$$

For the pH process, by differentiating (3) using (1) we get

$$\dot{y} = -c_y^{-1}(x,y)c_x(y)[f(x) + g(x)u + p(x)d] \tag{8}$$

where

$$c_x = \frac{\partial c(x,y)}{\partial x} = [c_{x1} \quad c_{x2}],$$

c_{x1}=1,

$$c_{x2} = \frac{1+2\times 10^{y-pk_2}}{1+10^{pk_1-y}+10^{y-pk_2}}.$$

$$c_y = (\ln 10) \times \left\{ x_2 \left(\frac{10^{pk_1-y} + 10^{y-pk_2} + 4 \times 10^{pk_1-pk_2}}{(1 + 10^{pk_1-y} + 10^{y-pk_2})^2} \right) + 10^{y-14} + 10^{-y} \right\},$$

Since $-c_y^{-1}(x,y)c_x(y)g(x) \neq 0$ for all x and y of interest, the relative degree $r = 1$, and the linearization relation is

$$u = \frac{v + [f(x) + p(x)d]c_y^{-1}(x,y)c_x(y)}{-c_y^{-1}(x,y)c_x(y)g(x)}$$

Now, assume v is the output of a PID controller in the form

$$v = k_P e + k_I \int e + \dot{y_{sp}} \tag{9}$$

where $e = y_{sp} - y$ and y_{sp} is the reference signal, k_P is the proportional gain, k_i is the integral gain.

The control structure of the FL controller is shown in Figure 3.

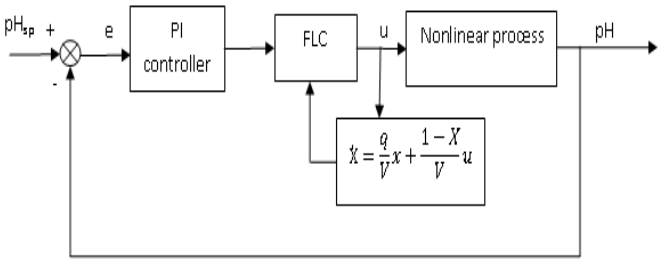


Figure 3: The control structure of the FL controller

4. ADRC Solution for the pH Process

It is shown that the control structure using the feedback linearization method is very complicated. To simplify the control design, ADRC is designed. It is shown that all the nonlinear and time-varying terms under ADRC are components of the total disturbance that need to be estimated and rejected, making the design and structure much more straightforward.

4.1. Conventional linear ADRC

We can rewrite the system described in (1) as

$$\dot{y} = f + bu \tag{10}$$

where f is the total disturbance, and b is the coefficient of the output change rate to the control variable step and varies all through the process. Now, ADRC task is to estimate f in real time and cancel it in the control law, which putting the extended state observer as a choice tool from the modern control theory. Define

the cases as $x_1 = y$ and $x_2 = f$, where the extended state is represented by x_2; then, Eq. (10) rewritten as

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} + \begin{bmatrix}b\\ 0\end{bmatrix}\dot{f}$$

$$y = x_1 \tag{11}$$

An ESO can be constructed to estimate the states

$$\begin{bmatrix}\dot{z}_1\\ \dot{z}_2\end{bmatrix} = \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}\begin{bmatrix}z_1\\ z_2\end{bmatrix} + \begin{bmatrix}b\\ 0\end{bmatrix}u + \begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix}(y - \hat{y})$$

$$\hat{y} = z_1 \tag{12}$$

where β_1 and β_2 represent the gains of the observer.

The total disturbance $z_2$will be estimated by a well-tuned observer in Eq.(12), and the control law

$$u = \frac{-z_2 + u_0}{b} \tag{13}$$

where u_0 is a virtual control signal, then Eq.(10) reduced to

$$\dot{y} = u_0 + (f - z_2) \approx u_0 \tag{14}$$

which can be easily controlled using a proportional controller.

$$u_0 = k_p(r - z_1) \tag{15}$$

Here k_p is the controller gain and r is the reference signal.

The structure of the ADRC is shown in Figure 4.

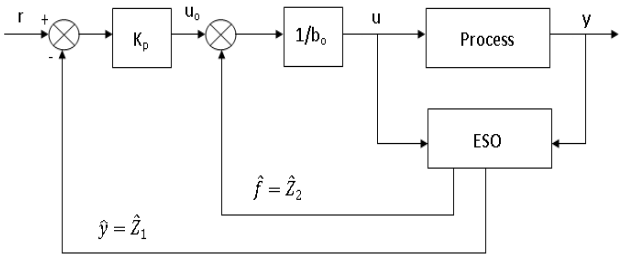


Figure 4: ADRC structure for a first-order process

The gains of the observer are chosen as $\beta_1 = 2\omega_{ob}$ and $\beta_2 = \omega^2_{ob}$ to put the observer poles at $-\omega_{ob}$. In the same way, the controller gains are selected as $k_p = \omega_c$ to put the pole of the control loop at $-\omega_c$. ω_{ob} and ω_c are called the observer bandwidth and controller bandwidth, respectively.

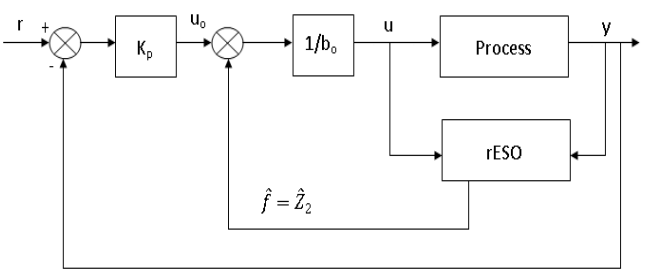


Figure 5: ESO based reduced-order ADRC

4.2. Reduced-order ADRC (rADRC)

In the ADRC, a second-order ESO is used to obtain the extended state compensation estimate; however, redundancy occurs when the state's first component can be directly measured [31]. A reduced-order state observer is employed, as shown in Figure 5.

The state observer of reduced-order for x_2 is

$$\dot{z}_2 = -\omega_{ob}z_2 - \omega_{ob}bu + \omega_{ob}\dot{y} \tag{16}$$

where ω_{ob} (>0) determines the convergent rate of the ESO. To avoid intensifying measurement noises by direct numerical differentiation of signal y in (16), we define a new state as

$$z = z_2 - \omega_{ob}y \tag{17}$$

Therefore (16) may be rewritten as

$$\dot{z} = -\omega_{ob}z - \omega_{ob}{}^2 y - \omega_{ob}bu \tag{18}$$

which combined with

$$z_2 = z + \omega_{ob}y \tag{19}$$

constitutes the reduced-order ESO. The remaining controller design is the same as those above.

5. Simulation Results

A comparison between the performances of the designed FL and ADRC controllers is carried out through simulation. The change in buffer and feed stream flow rate is regarded as a disturbance to the system and shown in Fig. 6. The ADRC parameters are shown in Table 2.

Table 2. ADRC controller parameters

ADRC	$\omega_{ob} = 5$	b = 0.05	$\omega_c = 10$
rADRC	$\omega_{ob} = 0.9$	b = 0.05	$\omega_c = 10$

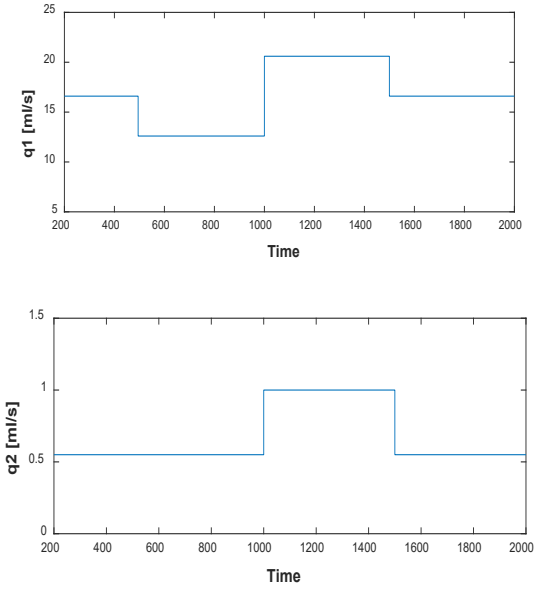


Figure 6: The change of buffer and feed stream flow rate

There are two cases considered in this simulation.

A. Constant Parameters

First, let us look at the case of the constant parameter. The efficiency of both disturbance rejection (Figure 7) and setpoint tracking (Figure 8) is examined to evaluate the effectiveness and robustness of the control system. A comparison to the FL controller carried out to illustrate the advantages of the ADRC as well as the rADRC. It can be observed from Figure 7 the superiority of ADRC and rADRC over the FL controller. We can see from Figure 8 that ADRC tracks the change in setpoint very well, and rADRC tracks it even better, while the FL controller shows some overshoot with longer settling time.

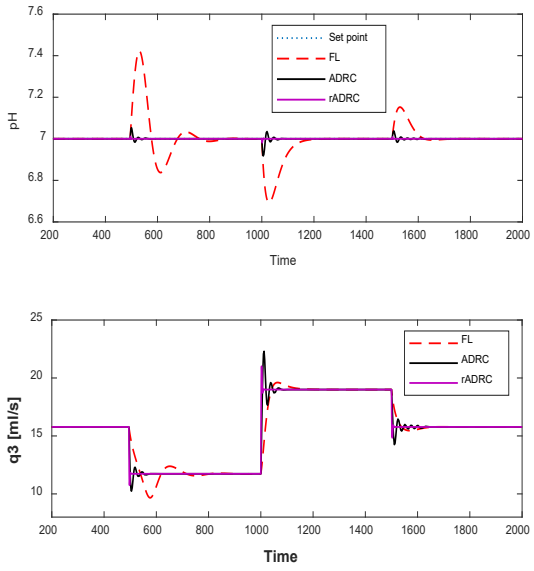


Figure 7: The performance of disturbance rejection

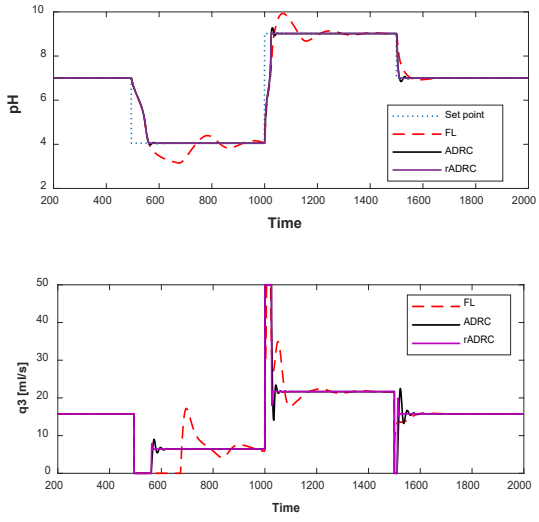


Figure 8: The performance of setpoint tracking

B. Variations in the parameters

The parameters changes are discussed In this case. They are regarded and handled as a system's internal disturbance. Figure 9 shows the changes in the concentrations of the feed stream.

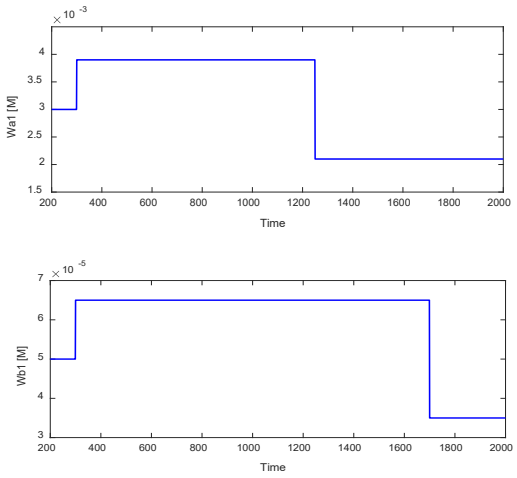


Figure 9: Variations in the concentration of the feed stream

Figure 10 shows the performance of the disturbance rejection, while Figure 11 shows the performance of the setpoint tracking. We can see obviously that ADRC and rADRC are superior to the FL controller.

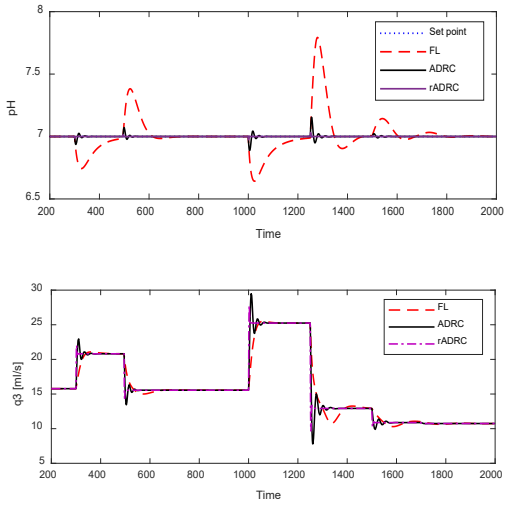


Figure 10: The performance of disturbance rejection with variation in parameters

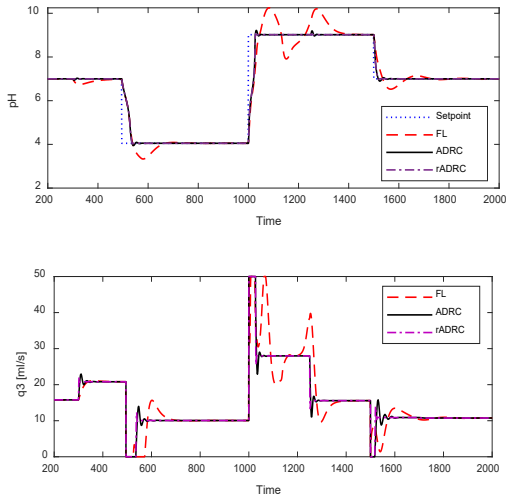


Figure 11: The performance of setpoint tracking with variation in parameters

6. Conclusion

A novel control technique (ADRC) for a pH process is presented in this paper. ADRC is an excellent solution to the process that needs a high performance controller to overcome the modeling errors and disturbances. ADRC's pH simulation studies showed a major enhancement over nonlinear FL control. For more analysis, a reduced-order ADRC was designed and proved that it has the most significant performance in regulating the pH in the existence of the input and external disturbances. The ADRC method is a strong algorithm, particularly in rejection of disturbances. Compared to FL control technique, design and tuning are much more straightforward.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] M. H. Eljack, M. B. Musa, W. Tan, "pH Neutralization Process Control Using Linear Active Disturbance Rejection Control." In 2018 IEEE 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP). 205-210, 2018. https://doi.org/10.1109/iccwamtip.2018.8632617

[2] M. A. Henson, D. E. Seborg, "Adaptive nonlinear control of a pH neutralization process," IEEE transactions on control systems technology, 2(3), 169-182, 1994. https://doi.org/10.1109/87.317975

[3] F. G. Shinskey, pH and PION Confro1 in Process and Waste Streams.New York: Wiley, 1973. https://doi.org/10.1016/0013-9327(74)90073-1

[4] G. K. McMillan, pH Control. Research Triangle Park, NC: Instrument Society of America, 1984.

[5] A. Johnson, "The control of fed-batch fermentation processes-A survey," Automatica, 23(6): 691-705, 1987. https://doi.org/10.1016/0005-1098(87)90026-4

[6] R. A. Aguiar, I. C. Franco, F. Leonardi, & F. Lima. Fractional PID Controller Applied to a Chemical Plant with Level and pH Control. Chemical Product and Process Modeling, **13**(4), 2018. https://doi.org/10.1515/cppm-2018-0004

[7] R. Vilanova, A. Visioli, PID Control in the Third Millennium: Lessons Learned and New Approaches", Springer, London 2012.

[8] R. A. Wright, C. Kravaris, "Nonlinear pH control in a CSTR," American Control Conference Proceeding, 1540-1544, 1989. https://doi.org/10.23919/acc.1989.4790431

[9] T. K. Gustafsson, K. V. Waller, "Dynamic modeling and reaction invariant control of pH," Chem. Eng. Sci. **38**(3): 389-398, 1983. https://doi.org/10.1016/0009-2509(83)80157-2

[10] P. Jutilia, "An application of adaptive pH-control algorithms based on physio-chemical modeling in a chemical waste-water treatment plant," Int. J. Control. **38**(3): 639-655,1983. https://doi.org/10.1080/00207178308933099

[11] G. Pajunen, "Comparison of linear and nonlinear adaptive control of pH-process," in Proc. IEEE Control Systems Magazine. **7**(1): 39-43, 1987. https://doi.org/10.1109/mcs.1987.1105238

[12] J. R. Panish, C. B. Brosilow, "Nonlinear inferential control," AIChE J., **34**(4): 633-644, 1988.

[13] W. C. Li and L. T. Biegler, "Newton-type controllers for constrained nonlinear processes with uncertainty," Ind. Eng. Chem. Res., **29**(8): 1647-1657, 1990. https://doi.org/10.1021/ie00104a012

[14] G. L. Williams, R. R. Rhinehart, J. B. Riggs, "In-line process model-based control of wastewater pH using dual base injection," Ind. Eng. Chem. Res., **29**(7): 1254-1259, 1990. https://doi.org/10.1021/ie00057a029

[15] J. J. E. Slotine, W. Li, Applied Nonlinear Control, 2nd ed., Prentice Hall, Inc., Upper Saddle River, NJ, 1991.

[16] M. H. E. Elameen, M. N. Karsiti, & R. Ibrahim, "Nonlinear model feedback linearization control strategy of a pH neutralization process," In 2014 5th International Conference on Intelligent and Advanced Systems (ICIAS) (pp. 1-5). IEEE, 20014. https://doi.org/10.1109/icias.2014.6869488

[17] S. F. Graebe, M. M. Seron, G. C. Goodwin, "Nonlinear tracking and input disturbance rejection with application to pH control," Journal of Process Control, 6(2-3): 195–202, 1996. https://doi.org/10.1016/0959-1524(96)86054-9

[18] H. Degachi, W. Chagra, & M. Ksouri. "Nonlinear model predictive control for pH neutralization process based on SOMA algorithm." INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS, **9**(1), 391-398, 2018. https://doi.org/10.14569/ijacsa.2018.090153

[19] A. W. Hermansson & S. Syafiie. "Model predictive control of pH neutralization processes: a review." Control Engineering Practice, 45, 98-109, 2015. https://doi.org/10.1016/j.conengprac.2015.09.005

[20] J. Chen, Z. Liu, Y. Peng, Q. Guo, "A New Control Method in pH Neutralization Process." Springer, Berlin, 511-519, 2013. https://doi.org/10.1007/978-3-642-38524-7_56

[21] A. Grancharova, J. Kocijan, & T. A. Johansen, "Explicit output-feedback nonlinear predictive control based on black-box models," Engineering Applications of Artificial Intelligence, **24**(2), 388-397, 2011. https://doi.org/10.1016/j.engappai.2010.10.009

[22] M. Heredia-Molinero, J. Sanchez-Prieto, J. Briongos, M. Palancar, "Feedback PID-like fuzzy controller for pH regulatory control near the equivalence point," J. Process Control, 2014, **24**(7): 1023–1037. https://doi.org/10.1016/j.jprocont.2014.05.006

[23] M. H. Dolatabadi, M. N. Monfared, A. Fakharian, "Robust H-infinity control for the pH neutralization process based on fuzzy models." In: Fuzzy Systems (IFSC), 13th Iranian Conference on, IEEE, 1–6, 2013. https://doi.org/10.1109/ifsc.2013.6675681

[24] J. Q. Han, "Auto-disturbances-rejection controller and its Application, Control and Decision," January, **13**(1): 19–23, 1998. (In Chinese).

[25] J. Q. Han, "From PID Technique to Active Disturbances Rejection Control Technique," Control Engineering of China, **9**(3): 13–18, March 2002. https://doi.org/10.1109/tie.2008.2011621

[26] J. Q. Han., "Active Disturbance Rejection Control Technique – the technique for estimating and compensating the uncertainties," National Defense Industry Press, 197-270, 2008.

[27] Z. Gao, "Scaling and parameterization based controller tuning," Proceedings of the American Control Conference, Denver, **6**: 4989-4996, 2003. https://doi.org/10.1109/acc.2003.1242516

[28] Z. Chen, Y. Wang, M. Sun, & Q. Sun, "Convergence and stability analysis of active disturbance rejection control for first-order nonlinear dynamic systems," Transactions of the Institute of Measurement and Control, **41**(7), 2064-2076, 2018. https://doi.org/10.1177/0142331218794812

[29] J. Ping & B. Wang. "Development of User-defined Block Library for Active-Disturbance-Rejection-Control," In 2010 Third International Conference on Information and Computing, Vol. 2, pp. 151-154, 2010. IEEE. https://doi.org/10.1109/icic.2010.132

[30] K. V. Walker, P. M. Makila, "Chemical reaction invariants and variants and their use in reactor modelling, simulation, and control," Ind. Eng. Chem. Process Des. Dev., **20**(1): 1-11, 1981. https://doi.org/10.1021/i200012a001

[31] R. Yang, M. Sun, Z. Chen, "Active disturbance rejection control on the first-order plant." Journal of Systems Engineering and Electronics, **22**(1): 95-102, 2011. https://doi.org/10.3969/j.issn.1004-4132.2011.01.012

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 110-114 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Forecasting Bio-economic Effects in the Milk Production based on the Potential of Animals for Productivity and Viability

Gennadii Georgievich Cherepanov*, Anatolii Ivanovich Mikhalskii, Zhanna Anatolievna Novosrltseva

Department of Incomplete Data Control, V.A. Trapeznikov Institute of Control Sciences RAS, 117997, Russia

ARTICLE INFO

Article history:
Received: 07 May, 2019
Accepted: 05 August, 2019
Online: 03 September, 2019

Keywords:
Dairy Cattle Breeding
Profitability of Milk Production
Cow's Productivity and Viability
Forecasting Algorithms
Planning in Dairy Industry

ABSTRACT

The most important biological factors, mainly determining the economic efficiency of milk production, are the productivity potential and the level of viability of cows. The aim of the work is to predict the bio-economic effects in a heterogeneous population of dairy cows taking into account the decrease in the length of productive life with increase in milk productivity. In the considered model situation, with an increase in the productivity of cows, the total profit per cow per year increases, while the profitability of production may decline. The developed algorithms are supposed to be used for solving the problems of optimizing the management of the dairy cow herd and planning technological development in animal husbandry.

1. Introduction

Used abbreviations and terms: PL is length of productive life (the number of consecutive lactations of 305 days each); PP is the potential of productivity, VP is the viability potential, CR is constitutive resistance, dry period is the period from the end of lactation to calving; calving period is lactation period + dry period.

Due to the reduction in the total size of dairy cattle population, the rate of decline in milk production in the Russian Federation is not compensated by a positive trend of increasing the average productivity of cows. The use of imported breeding material on a large scale and the use of intensive industrial technologies in milk production are accompanied by a decrease in the level of biodiversity, an increase in "productivity diseases", a decline in reproduction, a deterioration in the quality and safety of the products obtained. Similar problems in the field of dairy cattle husbandry have arisen in recent decades in other countries, since the high productivity of cows is often associated with an increased frequency of culling. [1, 2]. Therefore, to ensure profitability and overall sustainability of milk production, it is necessary to intensify research in the field of biology of the viability and productive longevity of cows [2 - 4]. It is necessary to search for biological traits and technical tools for the collection and transmission of information for early prediction of the health status of lactating cows at the earliest possible stages of postnatal life [2, 3].

Previously, it was shown that for early assessment and prediction of viability in groups of dairy cows, the trait of constitutive resistance can be used as an integral indicator reflecting the effects of age-related decline in protective forces [5], as well as the reciprocal of the rate of cows culling in first lactation. An analysis of empirical data reveals the presence of heterogeneity in terms of survival parameters in populations of various animal and human species. The data from bio-medical studies show that the patterns of mortality in a population considered as one whole may differ significantly from the mortality frequencies in the constituent subpopulations [6, 7]. There are data indicating the heterogeneity of dairy cattle populations on survival parameters [8, 9]. In the context of the stated approach, the main factors determining the lifetime productivity of cows are PP and VP. Age dynamics of milk production in groups of cows with different PL is determined by a specific combination of these factors.

From the viewpoint of planning the development of the dairy industry, it is important to have methodological tools for a comprehensive assessment and forecasting the influence of the most critical factors in the selection, reproduction and housing condition of dairy livestock. In this paper, as a starting point in this direction, we propose a pilot model for predicting bio-economic effects when varying a set of factors, including milk

*Corresponding Author: Gennadii G. Cherepanov, Institute of Animal Physiology, Biochemistry and Nutrition, Borovsk, Kaluga region, 249013, Russia, Email: 8961242110@mail.ru

production potential and the rates of disposal of the herd in a heterogeneous population of dairy cows. When setting the initial data, an empirically fact of a decrease in PL with increasing milk production was taken into account, especially pronounced in highly productive cows.

The aim of this work is the systematization of previously identified empirical patterns and the development of a pilot numerical model for predicting the bio-economic effects in a dairy cows population which is non uniform in terms of PP and VP.

2. Determination of The Numerical Values of Parameters using Empirical Data

2.1. Milk Production Potential

To estimate the potential of milk production in the *i*-th homogeneous group, cumulative 305 d milk yield data are used for consecutive lactations, $y_{mi}(t)$, ranging from 4 ($t = 4$) to the last completed lactation t_{imax}. Age dynamics of 305 d milk yield is described by a three-component multiplicative function

$$y_{mi}=A_i \times\exp(-\exp(-bt))\times D^t \quad (1)$$

in which A_i is a potential maximum lactation yield (productivity potential, PP), the exp(-exp(-bt)) function describes an age-dependent increase in the ability to form milk with a gradual reaching a plateau level, the function D^t (constitutive resistance, CR, D<1) describes an age-related decrease in the ability to maintain lactation activity in the *i*-th group. The maximum yield is always less than A by an amount depending on the rate of decline of the CR. At small values of D, the peak of milk yield for lactation occurs earlier and the cows drop out of the herd more quickly.

Earlier, when analyzing empirical data on cows of the Kholmogor and Black-and-White breeds, a linear positive relationship was found between PL, expressed by the number of last lactations t_{imax}, and the value of D_i in groups [5]. Thus, the value of D has the meaning of viability potential (VP), estimated from the age dynamics of 305 d yields in groups. The value of the parameter b in the second component of formula (1) varies within small limits (0.4–0.5) and for proximate calculations it was assumed to be 0.45.

2.2. *Potential of Viability*

The presence of a positive relationship between PL and the "initial" value of the CR in the analysis of the age dynamics of milk yield agrees qualitatively with the identified relationship between PL and the viability index estimated from the dynamics of cows leaving the head (culling by sum of reasons). Culling rate is described by the Gompertz function:

$$y_c(t)=\Delta S(t)/(S(t)\times\Delta t)=B\times\exp(ct) \quad (2)$$

where $S(t)$ is the cohort size at the moment of time t, $\Delta S\ (t)$ is the reduction of the cohort size for the period of time Δt. For lactating cows, time is usually expressed as the number of the current lactation and $\Delta t = 1$. Parameters B and c can be determined not in cohorts (in groups of animals born during the same period of time), but according to the data of age structure of the herd (cross-sectional study).

The mean PL (T, the average number of lactations in a cohort) can be estimated from empirical data using an analytical expression [10]:

$$T=\exp(B/c)\times\mathrm{Ei}(B/c)/c \quad (3)$$

where Ei(.) is a special function (integral exponent).

For use in zootechnic practice, we can use the truncated Gompertz distribution, which does not take into account a small number of longevity record holders. Our numerical simulation showed that in the range of values B = 0.04–0.3, c = 0.2–0.8, the maximum PL value is approximately 2 times larger than the average PL i.e. $t_{max}\approx 2\ T$ or $T = t_{max}/2$ [8].

The value of the initial rate of disposal y_{c1} in formula (2) means the difference between the number of cows in the first and second lactation divided by their number in the first lactation. When analyzing production data on the age structure of the dairy herd, in particular, on five breeds of the USA and on eight farms of the Leningrad region, the existence of a positive linear dependence of T on the value of $1/y_{c1}$ was revealed [5]. Therefore, the value of $1/y_{c1}$ has the meaning of predictor of longevity, as measured by the dynamics of the disposal of cows from the population, herd or cohort [8].

The function, inverse to the intensity of disposal, is characterized by a decrease in the viability with age ("resistance to mortality" according to Gompertz). Thus, at t = 1, the "initial" disposal intensity $y_{c1} = B\ e^c$ is determined, and the reciprocal, $1/y_{c1} = B^{-1}\ e^{-c}$, has the meaning of the initial viability level.

3. Forecasting Milk Production Efficiency

3.1. Baseline Data for The Forecast

In the model situations under consideration, there are four groups with consecutive values A = 24, 20, 16, 12 thousand kg (the average for the lifetime milk yield is 12.4; 11.0; 9.35; 7.64 thousand kg); T = 2, 3, 4. 5 (t_{max} = 4, 6, 8, 10 lactations); D = 0.95; 0.91; 0.89, 0.88. An additional list of source data may include a large number of indicators. In this case, this list includes the calving period 13 months and the main cost indicators, including feed costs, the price of sold milk and culled cows (delivery at a meat processing plant, utilization of carcasses of sick cows, etc.).

3.2. Intermediate Calculations and Target Forecast

The initial data for each group are calculated:
- the average lifetime cumulative 305 d milk yield over 10 completed lactations (y_m);
- the number of replacements of the culled cow on the heifer for the total number of 10 lactations ($Nr = 10/T$);
- calendar term corresponding to 10 lactations (years);
- revenue from sold milk over 10 lactations with price of 30 rubles per 1 kg of milk, thousand rubles (milk income, $m.inc = 30 \times ym$);
- revenue from sold culled cow over 10 lactations, thousand rubles (realization income, $real.inc = cow\ price \times Nr$);
- total revenue, thousand rubles (*income, inc = inc.m+ real.inc*);

- the average price of grown or purchased heifers, thousand rubles (heifer cost, *Hc*);
- the cost of growing or purchasing heifers, thousand rubles (reproduction expenses, *repr.exp.* = *Hc* × *Nr*);
- the cost of maintenance and technological spending per cow for the whole period, thousand rubles (technological expenses, *t.exp.* = *12* × *year*);
- feed costs for milk production per 10 lactations, thousand rubles (feed expenses, *f.exp* = *0.6* × *inc.m*);
- the sum of expenses, thousand rubles (*s.exp* = *repr.exp* + *t.exp* + *f.exp*);
- profit, thousand rubles (income over expenses, *IOE* = *inc.);*
- profitability (% rentability, *rent*).

The adopted forecasting scheme assumes a certain method of combining in the same algorithm two heterogeneous data, i.e. the age dynamics of productivity and the longevity of cows. In the literature, there is no description of such an algorithm, and there are also no systematic empirical data necessary for detailed analysis. The simplified calculation method used in this work for ten 305 d lactations at one value of calving interval is as follows. Since the average daily milk yield values are used in groups, to obtain a forecast of the calendar period for 10 lactations per cow place, you can use the above ratio $T = t_{max}/2$. For example, at $T =$ 5 it is necessary to have two periods, i.e. 2 cows with 5 lactations and 4 dry periods (100 months of lactation, 24 months of dry period). At the same time, in the case of culling, a new primiparous cow is transferred to this place, one period of dry period is missed, and the total calendar period for which the cost of one cow per year is calculated for this group is 10.3 years. For the next three groups, these values are 10, 9.9, 9.5.

4. Results and discussion

The results of calculating the profitability of milk production for 4 model groups (subpopulations) with different values of the productivity potential, the average PL, the cost of raising or importing young stock and the main technological costs with an interval of 13 months are given in Table 1 and Figure 1 and Figure 2.

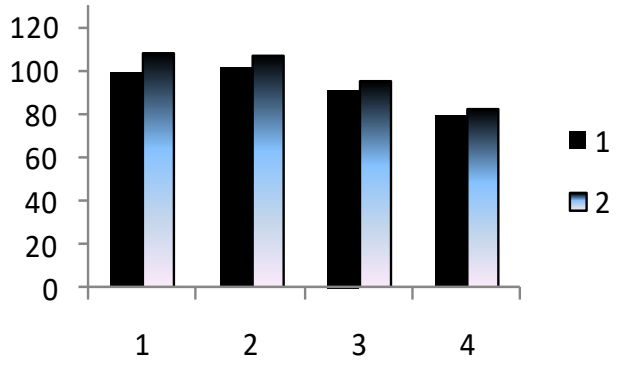


Figure. 1. Total profit per cow per year in the production of milk in four groups of cows with the values of the parameters listed in Table. In groups 1, 2, 3, 4, the average milk yield per life is 12.4; 11.0 9, 35, 7.64 thousand kg.; the calving period is 13 months. The average price of a culled cow sold in option 1 is 10, in option 2 it is 30 thousand rubles.

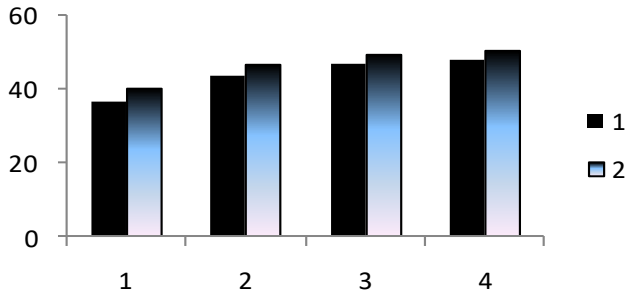


Figure. 2. Profitability of dairy raw materials production (revenue per milk and from sale of culled cows / sum of expenses) in four groups of cows with the values of the parameters listed in Table. In groups 1, 2, 3, 4, the average milk yield per life is 12.4; 11.0; 9.35; 7.64 thousand kg.; the calving period is 13 months. The average price of a culled cow sold in option 1 is 10, in option 2 it is 30 thousand rubles.

In the considered model situation, with an increase in the productivity of cows, the total profit per cow per year increases, while the rentability of production decreases.

To predict bio-economic effects in real conditions (including in terms of long-term planning), it is necessary to take into account the variability of interrelations and a wide range of possible values of parameters. In this work, seven factors were considered, the numerical value of which can vary widely, which is the case for different regions of the Russian Federation. At the same time, four gradations of two factors (level of productivity and viability index) were taken into account with two levels of one additional factor. At this stage of the study, the operability of the formulated concept and the pilot prediction model were tested. With a positive assessment of the results of the first stage, the expansion of the empirical base becomes urgent, i.e. obtaining a sufficient amount of production data while improving the calculation algorithm.

In the extended version of the model, additional indicators may be introduced into the calculations, including the reproduction parameters of the dairy herd. Since, at high levels of milk production, there is a shortage of the young animals number, to preserve the livestock population size the complex methods are developed (separation of bull-sires semen, induction of superovulation, *in vitro* fertilization, transplantation and embryo cloning), which increase the cost of reproduction of the dairy herd. To provide enhanced herds reproduction, the choice of the most appropriate method depends largely on the conditions of the territorial distribution of specific farms, as the initial data in the forecasting can vary greatly for small farms in the regions and for large industrial complexes with intensive technologies located near large urban agglomerations. Therefore, in order to rationalize management decisions, multivariate model calculations are required.

Researchers working in the field of studying the dynamics of human populations have identified the need to identify subpopulations in order to correctly predict demographic processes [11, 12]. In the process of breeding work in animal husbandry, similar problems arise, since an early assessment of the breeding value of parents for productive longevity of offspring is important. For bull-sires, one of the options for such an assessment may be

Table. Baseline data and results of the prediction of the bioeconomic efficiency of milk production in groups that differ in productivity potential (A) and viability (D).

D	*T*	*A*	*ym*	*Nr*	*Hc*	*years*	*repr.*	*t.exp.*	*inc.m*	*s.exp./year*	*inc./year*	*rent.*
0,95	2	24	124	5	80	10,3	400	124	969,2	268	98,9	37
0,91	3	20	110	3,3	65	10	214	120	980,7	231	101,4	44
0,89	4	16	93,5	2,5	50	9,9	125	119	878,2	195	91,2	47
0,88	5	12	76,4	2	35	9,5	70	114	732,8	164	79,2	48

Notes: notations are given in the section 3.2.

the inverse of the rate of disposal in first lactation, measured in a group of daughters [13]. When using the latest genomic, bioinformatics and telecommunication technologies, there are ample opportunities for solving such problems. The use of microelectronic sensors and wireless communication allows to increase the amount of information captured during screening surveys throughout the animal's life, which can dramatically increase the possibilities for monitoring and forecasting multifactorial bio-economic effects.

At this stage, it is important to compare the initial prerequisites of our calculations with known ideas and concepts related to the problems of populations viability. The concept of constitutive resistance as the main determinant of productive animal longevity, used in this work, is in many ways consistent with the concepts of creod according to Waddington [14] and primal health of Odent [15]. There is evidence that metabolic diseases occurring in later life may arise *in utero* as a result of suboptimal intrauterine conditions. The process by which prenatal defects lead to permanent changes in tissue structure and function, and finally to low birthweight, is known as developmental programming; i.e. placental insufficiency invariably affects embryonic development and health in later life [16]. So, the level of viability in heifers is being formed epigenetically in the prenatal and postnatal periods when exposed to internal and external factors. Our data indicate that the value of cow PL depends on the "initial" level of viability indicators (at the beginning of the reproductive period). The existence of this dependence is indicated by empirical data and, as carried out by calculations, a similar relationship follows from the general properties of the Gompertz function, traditionally used in analyzing the survival of populations.

5. Conclusion

The most important biological factors, mainly determining the economic efficiency of milk production in dairy farms, are the productivity potential and the level of cow viability formed to the age of reproductive activity. The length of the productive life of cows largely depends on the initial level of viability indicators. Currently, the average group potential levels of productivity and viability can be determined from indirect information – the data of age dynamics of 305 d milk yields and intensity of cows culling in first lactation.

From the viewpoint of planning the development of the dairy industry, it is important to have methodological tools to predict the influence of the most critical factors in the selection, reproduction and housing condition of dairy livestock. In this paper, we tested the operability of the general approach and the algorithm of numerical simulation using discrete scales and data approximation by the Gompertz function. The results showed that in the considered model situation, with the increase of the average of lifetime milk yield, the total profit per cow per year increases, while the profitability may decline. According to the authors, there is reason to assume that the identified consistent patterns reflect a general trend, at least in a qualitative sense, since empirically established facts and regularities are taken into account in the set of initial data.

After expanding the volume of empirical data, setting parameters and modification of algorithm taking into account the features of real objects, the proposed forecasting model is supposed to be used to optimize the herd management and forecast scientific and technological development in dairy cattle husbandry.

References

[1] K. Adamczyk, J. Makulska, W. Jagusiak and A. Węglarz, "Associations between strain, herd size, age at first calving, culling reason and lifetime performance characteristics in Holstein-Friesian cows" Animal. 11(2): 327-334, 2017. DOI: 10.1017/S1751731116001348.

[2] F. Miglior, A. Fleming, F. Malchiodi, L.F. Brito, P. Martin and C.F. Baes, "A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle" J. Dairy Sci. 100(12): 10251-10271, 2017. DOI: 10.3168/jds.2017-12968.

[3] E. Vasseur, "Animal behavior and well-being symposium: optimizing outcome measures of welfare in dairy cattle assessment" J. Anim. Sci. 95(3):1365-1371, 2017. DOI: 10.2527/jas.2016.0880.

[4] J.H. Britt, R.A. Cushman, C.D. Dechow, H. Dobson, P. Humblot, M.F. Hutjens, G.A. Jones, P.S. Ruegg, I.M. Sheldon and J.S. Stevenson, "Invited review: Learning from the future-A vision for dairy farms and cows in 2067." J. Dairy Sci. 101(5): 3722-3741, 2018. DOI:10.3168/jds.2017-14025

[5] G.G. Cherepanov, "Justification of the concept of the key role of constitutive resistance for the viability and duration of the use of highly productive animals" Problems of Productive Animal Biology, 4: 5-34, 2014 [In Russian]

[6] J.W. Vaupel, J.R. Carey, K. Christensen, T. E. Johnson, A. Yashin, N.V. Holm, I.A. Iachine, V. Kannisto, A.A. Khazaeli, P. Liedo, V.D. Longo, Y.Zeng, K.G. Manton and J.W. Curtsinger, "Biodemografic trajectories of longevity" Science, 280: 855-860, 1998.

[7] R.V. Ivanov, A.I. Mikhalskii, V.K. Ivanov, S.Yu. Chekin, M.A. Maksyutov and V V. Kashcheev, "On identification of morbidity parameters in a heterogeneous model: The cases of complete and incomplete information" Autom. Remote Control, 78(7): 1329-1340, 2017. DOI: 10.1134/S0005117917070141

[8] G.G. Cherepanov, A.I. Mikhalsky and Zh.A. Novoseltseva, "Estimation of survival parameters for components of a heterogeneous population of productive animals: problem analysis, approximate solution options" Problems of Productive Animal Biology. 4: 83-97, 2017 [In Russian]

[9] G.G. Cherepanov, A.I. Mikhalskii and J.A. Novoseltseva, "Estimation of survival parameters in heterogeneous population" 11th IEEE International Conference on Information and Communication Technologies (AICT). Moscow, Russia, 2017: 433-435. [In Russian]

[10] A.V. Krementsova and N.V. Gorbunova, "The role of environment in the dynamics of the life expectancy distribution". Autom. Remout Control 8: 121-133, 2010.

[11] N. Keyfitz and G. Littman, "Mortality in a heterogenous population" Population Studies, 33: 333-342, 1979.

[12] A.I. Mikhalskii, "Methods of analysis of heterogenic structures in population" Moscow: IPU RAS Publ., 2002. [In Russian].
[13] G.G. Cherepanov, "Prediction of viability of cows: a new look at the old problem" Agricultural Research and Technology. Open Journal (ARTOAJ). 141(5), 2018. DOI: 10.19080 / ARTOAJ.2018.14.555931
[14] C.H. Waddington, "Basic biological conceptions". In: *Na puti k teoreticheskoi biologii. I. Prolegomeny*. Moscow, Mir Publ, P.11-38. 1970. Translated from English: Towards a theoretical biology. I. Prolegomena. IUBS Symposium (Ed. C.H. Waddington), Birmingham: Aldine Publ., 1968.
[15] M. Odent, "Primal Health" London, Century Hutchinson, 1986.
[16] G. Opsomer, M. Van Eetvelde, M. Kamal and A. Van Soom. Epidemiological evidence for metabolic programming in dairy cattle. Reprod. Fertil. 29(1): 52-57, 2016. DOI: 10.1071/RD16410.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 115-122 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Survey on Semantic Similarity Based on Document Clustering

Rowaida Khalil Ibrahim[*,1], Subhi Rafeeq Mohammed Zeebaree[2], Karwan Fahmi Sami Jacksi[1]

[1]*Computer Science Department, University of Zakho, 420011, Zakho, Iraq*

[2]*Department of IT, Akre Technical College, Duhok Polytechnic University, 42004, Akre, Iraq*

ARTICLEINFO

Article history:
Received: 15 June, 2019
Accepted: 16 July, 2019
Online: 03 September, 2019

Keywords:
Clustering
Text Clustering
Semantic Similarity
Ontology
Evaluation Measures

ABSTRACT

Clustering is a branch of data mining which involves grouping similar data in a collection known as cluster. Clustering can be used in many fields, one of the important applications is the intelligent text clustering. Text clustering in traditional algorithms was collecting documents based on keyword matching, this means that the documents were clustered without having any descriptive notions. Hence, non-similar documents were collected in the same cluster. The key solution for this problem is to cluster documents based on semantic similarity, where the documents are clustered based on the meaning and not keywords. In this research, fifty papers which use semantic similarity in different fields have been reviewed, thirteen of them that are using semantic similarity based on document clustering in five recent years have been selected for a deep study. A comprehensive literature review for all the selected papers is stated. A comparison regarding their algorithms, used tools, and evaluation methods is given. Finally, an intensive discussion comparing the works is presented.

1. Introduction

Clustering is a problem of unsupervised learning; the main task is to group a set of objects in a manner that the group of objects in the same collection called the cluster are more similar (in meaning) to each other than those in the other cluster [1][2]. Document clustering or (text grouping) is a block analysis application on text documents, has applications in automatic organization of document, filtering or fast information retrieval and subject extraction [3][4]. The text group divides a set of text documents into different categories so that the documents in the same category group describe the same subject. Text clustering is an essential function in the text mining process [5]. Problem of Description clustering is one of the key issues for the traditional algorithm of document clustering. The objects can cluster in traditional algorithm, but the result of clustering cannot give description concept. It is therefore necessary to find a way to resolve the problem of grouping the document description in such a way to meet the specific needs of document clustering [6]–[8].

For conceptual and meaningful collecting of documents into groups (clusters), it is necessary to exploit the semantic relationships between words and concept in the documents [9]. For this case semantic similarity is needed. The term semantic similarity is a specific measure for several pamphlets/concepts in contrast with the similarity that capable of roughly- calculations with respect to represent grammar (such as the format of the string). For this reason, the mathematical instruments are depended for asset estimation semantic association among language units and ideas or examples, by means of a numerical description obtained by comparing information that supports its meaning or description of its nature [10]–[12].

It is often confused with the term semantic relatedness and semantic similarity. The semantic relationship involves any relationship between two periods, whereas the semantic similarity only includes a relationship “is a” [13]. For example, "car" and “bus" are similar, but is related to “driving" and “road”. By determining topological similarity, Semantic similarity can be estimated using the ontology to determine the distance between concepts/term [14][15].

For semantic similarity purpose between concepts many tools are used, the most common one is the WordNet. The term of WordNet Is the lexicons of the English language database. It’s collecting the words into a group of synonyms named syn-sets, provides general, short info and register a different semantic relationship between these sets of synonyms. Especially WordNet is suitable for similarity procedures, because it regulates nouns and

[*]Rowaida Khalil Ibrahim, Email: rowayda.gravi@gmail.com

verbs in the hierarchy of "is a" relation [16]. Several tools for finding semantic similarity between concepts and words have been proposed in our literature but most of them have been used WordNet. For this paper we reviewed many papers that used semantic similarity but we choose only thirteen systems in these five recent years that using semantic similarity for clustering propose and we give some important information of each of them. Our paper organized as follow: in section 2 it's a literature review of papers, in section 3 we give table of discussion of all that systems we reviewed before, then we conclude all work in section 4.

2. Literature Review

The main framework of the approach addressed by [17] consists of three modules, first one is Page Counts based measures, second Snippets and third is taxonomy-based approach. for page count four co-occurrence standard measures are used namely (WebOverlap, WebJaccard, WebPMI and WebDice), after the page count of two words are calculating then checking the context of those two words locally by using snippets based on probability measure to find out the pattern exist between both words. Also, they have used taxonomy-based measure (LEACOCK and CHODOROW (LCH) [18], WU & PALMER (WUP) [19], and RESNIK [20]) that using WordNet for finding and computing similarity between words (X and Y). then for combination the result using tow class of Support vector Machin (SVM and libSVM). The term of dataset used for evaluation, for training propose using MEN dataset [21] Selecting synonymous word pair at random from dataset and create non-synonymous for each one. However, for testing propose using MC approach [22], and RG [23], WordSimilarity-353 (WS) [24]). And using WEKA [25] for testing and training of the SVM with suitable demonstrating approach LibSVM and the qualified prototypical via determining the division ratio is about 70.51%. The results of accuracy were obtained near by 72%.

As explained in [26], the approach in this paper at first finding similarity between tow documents then extend to between two document sets, after that using clustering and classification to group the documents based on similarity measures .the proposed measure of similarity called (similarity measure for text processing) SMTP and the classification algorithm used are KNN based (SL-KNN) single label classification, KNN based (ML-KNN) multi label classification and for clustering algorithm using K-mean Clustering and (HAC) Hierarchical Agglomerative Clustering. The result checks the SMTP similarity measure's effectiveness that applying in text application (SL-KNN, ML-KNN, K-mean and HAC). Also, Compare SMTP performance with other five metrics, Cosine, Euclidean, IT-Sim, Extended Jaccard (EJ) and Pairwise-adaptive (Pairwise) by using different k value and using evaluating measures like (AC (Accuracy), EN (Entropy)). It shows that the similarity measure usefulness may rely on: 1) the format of feature, e.g., tf-idf or word count; 2) application of domains, e.g., image or text; and 3) classification or clustering algorithm. In addition, the term of datasets which are used for evaluating are three different datasets WebKB download from internet, Reuters-8 [27], and RCV1 [28]. Finally, in each case the result show the proposed measure SMTP have better performance.

In [29], the system approach for clustering the text document is an ensemble approach by using concept from Wikipedia. Two kind of clustering used, First approach is Partitional Clustering (lexical document clustering (LDC)) have three main phases are (Clustering term, finding documents lexical seed and Clustering text documents), in phase1 for the clustering purpose the fuzzy c-means algorithm is used to collecting the columns of matrix the (document-term matrix) hence the Fuzzy c-means collect the document-terms into k term clusters then in phase2 extract the documents representative, that are using as seeds in order for clustering all documents later, finally in phase3 based on those document seeds of each cluster term the centroid of document will be computed. Then the distances among the centroids and documents are used for clustering the documents. Second approach is Ensemble clustering ELSDC (Ensemble Lexical-Semantic Document Clustering) the main phases are (clustering term and topic key term selection, lexical seed documents extraction, find documents seeds, tagged and by using the consensus method collecting the documents). In phases 1 and 2, is same as of (LDC) but in phases 3 and 4, BOC and BOW are used for documents representation. The relevant concepts extract from Wikipedia by wikify module. The Naive Bayes classifier is the component of consensus method, by using documents in the same cluster in both clustering they are trained. The final cluster is generated after the remaining document are classified. However, the feature selection Var-TFIDF method are used for both approaches.

The experimental result has two rounds, in first round find comparison between LDA model and lexical document clustering (LDC) algorithm, the LDA's main idea is that the document can be rate as a distribution probabilistic on the underlying themes seen each subject as the terms of probability distribution. The C++ implementation are used for the LDA model; in this round It has been shown that the LDC can generate results similar to the LDA model on some real text data sets based on the representation of the duration document. In second round of experiments, find comparison between LDC and proposed ensemble algorithm (ELSDC). Show the benefits of incorporating Wikipedia concepts into document clustering. For evaluation proposed using Normalized Mutual Information (NMI) hence It measures the amount of information obtained from the categories given to a group of clusters, and has a maximum value of one clustering process when re-create the layers perfectly and the minimum it is zero. Also, for experiment propose eight datasets are used (20Newsgroups (News-sim3, News-rel3 and 20ng-whole), Reuters-21578, Classic4 is created from SMART data repository, WebKB, SMS Spam Collection and Cade is gathered from the content of Brazilian web pages).

As show in [30], extracting document feature and document vectorization to find similarity between Arabic webpages and showing the semantic annotation. The main steps in proposed approach are, in the first step Extraction of main class features using (Arabic VerbNet). The second one is document vectorization according to semantic feature by two solutions (semantic class probability distribution or Semantic class density) and in third step doing clustering and annotation. the model is work as follow the annotated document is take to be classified and all opinion word will be recognized by using semantic feature extraction and all the word will be aggregate to give semantic annotation to the document. The result gathered tow experiments, at the first one is

showing comparison between clustering only using K-mean (standard) and clustering using K-mean with semantic class density however, k-mean with semantic probability distribution and this is done by using evaluation measures (purity, MICD and DBI) hence the purity show that the clustering with vectorization signifies with high score, MICD tend that proposed model is out perform than standard k-mean and DBI show that proposed model appear to have smaller value than standard is mean that is better performance. The Runtime of standard k-mean is longer than the time consumed for tow solutions k-mean with vectorization. The second experiment showing annotation by using Mean score among cluster with both vectorization solution and Mean ratio among corpus (MRAC) also with two solution of vectorization. However, the datasets used for evaluation propose are collected corpus from the archives of online Arabic newspaper and VerbNet download from internet.

In [31], the approach of semantic clustering the text documents by using lexical chain and WordNet. The proposed approach in this paper are work as follow, at first the polysemy and synonymy are two main problem that effect the representation of text, this problem is solve by using WordNet based on Word sense disambiguation, after that introducing the lexical chain to capture the main theme of text and find the relationship between the word sense, then is show that the method can find a true number of clustering which is useful for finding the number of k that use in K-mean clustering algorithm. Finally, for experimental propose the paper show the comparison between all three (Base, DC, DCS) and with another system like ASG03 [32], LMJ10 [33] and CSF11 [34], hence Base is mean (all nouns) the system use all basic preprocessing techniques, i.e. term set extraction, stop words removing, stemming word and identification of nouns from term set but without performing WSD. However, DC is (disambiguated concepts) corresponding to the Base, but with preforming WSD and finally DCS (disambiguated core semantics) also corresponding to the DC, but adds the core semantics extraction process and is the proposed method.

In addition, in this paper for evaluation propose using Reuters-21578 corpus dataset and the evaluation measure using purity, entropy and F1-measure. Finally, the result show that in all experimental scheme the LMJ10 is the worst and the term of dimensionality reduction the feature account that derived from ASGO3 is lower than the number of base and the reduction of DSC is between 10% and 40% hence that the CSF11 greater than 74%, as well as the cluster quality obtained using the core semantic feature is better than using all nouns and using disambiguated concept (or at least comparable to) however the performance of using the disambiguation concepts is better than using all nouns, This suggests that the proposed disambiguation standard can solve the problem in a highly commendable and volatile manner, improve quality to a certain extent, and that the features of the semantic strings produced by the DCS approach do not only reduce the number of semantic concepts without losing much information, also adequately the main topic of the document that helps the clustering.

As [9], the model of clustering proposed in this paper incorporate Coreference resolution and exploit semantic relationship among the words by tackling polysemy and synonymy problem using WordNet and semantic similarity. The proposed approach consists of five modules are (Coreference Resolution, Preprocessing, Synonymy Identification and Sense disambiguation, Feature Selection and Bisecting k-means) modules. the purpose of those five model are, at first Coreference Resolution is the process of identifying Coreference words that occur in the document, then the propose of preprocessing model to transform the document in more suitable form and this is done in three steps (POS tagging ,Stop word Elimination and stemming),after that Sense Disambiguation and Synonym Identification deal with polysemy and synonymy problem in document and the model use WordNet for this propose, then the feature selection done by weighing the words in document using tf-idf, finally the documents are clustered by using Bisecting K-mean. For experimental propose the paper using four classic Datasets (CACM are 60 Documents, CISI are 44 Documents, CRAN are 44 Documents, and MED consists of 52 Documents) and for evaluating the quality of cluster using purity. The result show comparison between Base and proposed model hence the Base is the model without (the Coreference resolution, sense disambiguation and pruning term), after comparison the term of purity result show that the proposed model is observed that achieve 30% of improvement in clustering purity, that the purity rate of Base configuration is 0.55 hence the proposed model is 0.8.

As [35], Based on style Similarity and structure Clustering Web Pages, the model which used in this paper is DOM (Document Object Model) tree. The approach using two main measures, one of them for structural similarity to find similarity on DOM using TED (Tree Edit Distance) for HTML pages and another one using (Jaccard similarity measure) for CSS which called Stylistic similarity on style sheet information, after that based on those similarity measures using (Near neighbor clustering technique) for clustering propose, however to combine this information using Jarvis and Patrick method (shared nearest neighbors) [36]. It is a respectable feature aimed at aggregating Internet pages built for pairs similarity metrics. For evaluating propose using different threshold value of near neighbor at first using 90% edge for resemblance toward luxury near neighbors' pamphlets and 90% threshold for communal close towards combine clustering then altered the threshold to 95% until found the optimal threshold. The result show that when the threshold was 90% total number cluster was 12, and when it was 95% the total was 24 cluster as well as when it was 85 the total number of clusters was only 5, so after few trails they found that for test dataset the optimal threshold for both resemblances and intersection for communal close parties was near by 85%. As it declares that the result is positive as early evaluation.

As [37], the approach based on Hierarchical agglomerative clustering method (HAC) that is based of conceptual annotation of documents. The key steps of approach are: (1) pairwise calculation of the initial similarity matrix (2) Explanation of newly cluster and updating the matrix of similarity (3) post processing of HAC tree result. The approach chooses to rely on Lin Measure for pairwise similarity and on Base Match Average (BMA) for GroupWise semantic similarity, the proposed method uses GroupWise semantic similarity to compute the pairwise similarities between document for creating the label-base similarities matrix. When two clusters are agglomerative a new cluster are creating then automatically compute the annotation for this new cluster and then the label-based matrix are iteratively update, finally using

postprocessing of the tree hierarchy to take the branch lengths advantage in the tree to only keep the most meaningful cluster. For evaluation propose using (HSC heavy semantic clustering and LSC light semantic clustering) approaches for comparison with baseline (classical HAC with naive cluster annotation) by using Normalized Robinson-Foulds Distance measure hence the lower the distance the higher resemblance. The result after comparison between approaches (HSC, LSC and baseline) the HSC and LSC clearly outperform the baseline but for runtime the baseline was faster. then it shows the comparison of approaches (LSC, Baseline) with postprocessing and without post processing (pp and nopp) show that both are bad without postprocessing, finally it evaluating cluster labeling between our clustering labeling (CL), naive merge annotation (baseline) and expert also, the result show CL is better. And the dataset which used for evaluating propose was derived from WordNet-based disambiguation version of the aforementioned bookmark annotation [38] however, all resources and result of this paper are available at [39].

As [40], the approach for semantic document clustering based on the graph similarity. Graph is mainly used information from WordNet to the degree of semantic similarity between 150,000 of the most common in English terminology. The proposed approach at first Adds all the documents in to the similarity graph, then the distance between a pair of documents is measured by evaluating the paths between them, where a path in the graph can go through several terms that are semantically similar. The idea is to creating a node for each document and connecting this node to the graph, then to find the semantic similarity between the two documents will be measured by calculating the distance between the two nodes, and using three choices for the distance metric: the cosine, linear or logarithmic. Finally, for clustering propose using K-mean algorithm to cluster a set of documents and the algorithm relies on a way for computing the distance between two documents. The term of experimental result, the dataset of documents from Reuters-21578 benchmark and using F-measure for evaluating. The end of evaluating after applying the K-mean algorithm in three different distance metrics (cosine, linear and logarithmic) at first the result show that Using the similarity graph can lead to both higher precision and recall. And with Noting that using the linear or logarithmic similarity metric did not make a difference. The reason is that the two metrics apply different monotonic functions on the average of the sum of the forward and backward paths. Applying these monotonic functions has no effect on the ordering of the distances between nodes and on the clustering result.

In [41] paper the approach named (An Ontology-based and Domain Specific Clustering Methodology for Financial Documents). The main steps in approach are (preprocessing and feature extraction, sense word disambiguation, representation document and clustering). The approach works as follow: at step1 the preprocessing and data extraction is responsible to extract text from document in order to transform word in to more meaningful form and this is done in three steps (Stop Word Removal, Noun Extraction and Lemmatization). In step2 word sense disambiguation the correct sense for the noun are identified and this is done by using (WordNet ontology/database and an information content file) and in this study two disambiguation techniques are used which used different external information to remove ambiguity, used one technique (Brown Content File) which is the default for measuring Resnik and other technical used (file content financial information) that is proposed in this study. In step3 the document representation the features are represented as term frequency (TF or tf) vector to be prepare for clustering. Then, in step4 clustering is the final step which clusters the document vector by using algorithm and in this approach two kind of algorithms are studies (K-mean and sequential Information Bottleneck).

The dataset which used in this study was downloaded from EMMA [42] online repository using 446 documents, EMMA documents are classified under the US jurisdiction and their sectors or purposes and have achieved three types of labeling settings according to the categories provided by EMMA (setting1: purpose at the same time called sector as the class label, setting2: class label as state and setting3: class label as mixing of both state and sector). Then for evaluating propose in this study using purity for external evaluation to evaluate the cluster performance. The result showing the purity rate of both algorithm K-mean and sIB based on three settings also showing comparison as follow: 1) comparison between all term and noun only, Although the names are selected, the number of features is reduced significantly, the purity values of both algorithms do not have a considerable difference; 2) comparison between tf-idf of nouns and without tf-idf, it shows that with tf-idf better purity; 3) comparison between no disambiguation and with disambiguation, both methods of disambiguation have yielded good results, but the finest result produced with one the used financial information content file. Moreover, based on experimental results, the sIB algorithm can be determined as the most appropriate algorithm for aggregation.

As explained in [43], The system called WMDC (Wikipedia matching document classification) several steps are used in this approach. First, select knowledge and concept from Wikipedia. Second, using heuristic selection to pick up related concept. Third, finding similarity between document using combination of (Semantic similarity based on Wikipedia machine and Textual similarity based on keywords matching), and choosing K-mean algorithm for classification in this approach because of its efficiency and accuracy. Also, the evaluation experiment in this paper are divided in to two part. First part, focus on effectiveness of heuristic selection Rules, hence three rules are used to pick up the related concept of word (Rule1: all title, Rule2: all keyword and Rule3: any keyword) for evaluating efficiency of those three rules using selectivity measure and evaluating quality by using relevance. Second part, evaluating the effectiveness of approach.

Using Purity for measuring the accuracy of document classification and for evaluating the efficiency of approach using (vector construction time). The datasets used in this paper are Wikipedia dataset which download directly from internet and published in 2011-10-7, and document dataset from Reuters-21578 divided in to 82 cluster by removing those cluster with documents less than 15 and more than 200 only 30 cluster reminding to enhance the experimental effect. The result in this paper show that after using 4 different thresholds 0,0.5,0.10 and 0.15 for evaluating all rules of heuristic selection given that the smaller rule of selectivity is better efficiency and higher rule of relevance is better quality. then as the overall given that the proposed approach can accurately find out the related concept for a given document. Then the same threshold used for purity and (vector construction time) for evaluating classification accuracy and efficiency it concluded

that the proposed approach can improve the efficiency of language and for the corresponding Wikipedia under the precondition of not compromising the accuracy classification of the document.

As in [44] the system called SEMHYBODC related to (Fuzzy C-Manners). The methodology of clustering works as follow: At first the documents are annotated by "KIM" plugin before clustering. Then the documents are clustered using FPSOFCM (FPSO+FCM) and concept weight is calculated. The term of weight is recalculated through the steps specified in SEMHYBODC algorithm. And the Accuracy of clustering is computed using measures such as cluster purity and compared with various other hybrid approaches. after evaluating by using F-measure and purity the result show that, using swarm intelligence for clustering is not suitable for large dataset because the computational time is more but it give better accuracy in order to solve this problem a hybrid approach combine with algorithms such as PSO and FCM, however the PSO is combine with K-mean it's also leads to optimal solution but the time for computational is still higher, then as finally result the FPSO+FCM it show improvement over another algorithms like K-mean, Fuzzy C-means and Hybrid, It is combined with the ability to search from the globalized and fast algorithm PSO convergence algorithm FCM. It is used as a result of the initial seed FPSO algorithm FCM, which is applied to the purifying with generate concluding outcome. For evaluating propose webpages are collected from internet.

As explain in [45] many methods of clusters are operate based on similarity between documents as well as we explained before in our literature but in this paper the semantic similarity used for clustering the articles. The approach which proposed in this paper are at the first step the semantics of articles based on participating the entities in these articles, build three vectors of representations for each article: One calculates the average vector for all entities, one also with all entities but excluding citations and the other with only citation entities. Once the article vectors are generated, the next step is to identify clusters of articles based on vectorization using (k-mean) Clustering and using the Louvain (Network-based clustering methods) for community detection. For experiments the both clustering methods are using the Astro dataset [46]. At first four solutions of clustering are collected, namely CWTSC5, STS-RG, UMSI0 and ECOOM-BC13 for the pseudo-ground-truth based on adjustF1 that the best K are choosing which gives a highest score of adjustF1 also using adjustedF1 score to evaluate the results of clustering as well from the Louvain method.

The result shows the quality scores of three cluster based on pseudo-ground-truth (no citation, only citation, all entities) and OCLC Louvain then gives the average Adjusted Mutual Information scores (AMI) between current solution and the all four other solutions named, STS-RG, UMSI0, CWTS-C5 and ECOOM-BC13. It indicates that, if citations are used only, the resulting combinations correspond to other solutions of clustering than those if not using citation whose adjusted F1 score is also the lowest. Not surprisingly, the other clustering solutions depend heavily on citation information. So, even if they use different methods of citations, the citation information is still bringing enough agreement between them. Use all entities to represent articles has highest score adjustedF1 and agree with most others. Also, find AMI results among these three solutions and groups based on the Louvain method. Again, the cluster based on citation only agree with the results of Louvain almost as much as using cluster that use all entities. In accordance with these measures, they decided to use all entities as a major choice of characteristics, and to maintain 31 sets as key results for K-Means, which is named OCLC-31.

3. Discussion

The table below show the survey of approaches that using semantic similarity based on clustering. The table give a number of papers each paper has specific method and different tools are used for each of this method as well as provide information of each paper given different options like Measures for similarity and evaluation propose and algorithms for clustering or classification also, provide datasets that used by each paper.

As show in the survey, most of the method using WordNet, the term of WordNet as we mentioned before in introduction is the lexicons of the English language database, as well as in information system the WordNet was used for a number of proposals, for example (information retrieval, word sense disambiguation, text classification automatically, summarization of text automatically, machine translation and also can used for crossword puzzle generation automatically) [47] but one of the WordNet most common application is using to determine the similarity between words as we seen in all that method which are using WordNet the main propose was used for (word sense disambiguation) that deal with polysemy and synonymy problem. However, rather than WordNet there are another tool are used like Wikipedia or Kim Plugging also, there are method based on clustering, similarity graph or even network. Also, most of people using K-mean algorithm for clustering hence it's a must popular clustering algorithm that grouping the documents using a similarity metric that is based on keywords matching, and its useful for huge variables for this reason most of time computationally the k-mean is faster than hierarchical clustering but if keeping (K) small, and another advantage is it can make a tighter cluster [48]. However, many of research using bisecting K-mean because it can deal with large dataset also its very satisfactory quality of clustering with low cost. And another algorithm used as shown in surveys like fuzzy c-mean, HAC, Naïve base classifier, near neighbor, sIB, Louvian method and using PSO with clustering algorithm for optimization propose.

As it can be seen that other measures are used like tf-idf its mean (term frequency-inverse document frequency) it's using as weighting factor the main job is determine how much important the word is to the document in corpus or in group [49], as we seen in survey most of paper used this measure for feature selection. Also, other measures are used like Var-TFIDF, heuristic selection etc. for evaluation and experimental result many types of dataset are used and there are specific measures that using for evaluation purpose most popular one is Purity hence its very simple and primary evaluation measure that using for validation to determine the quality of cluster [50], however there are another one like F-measure, Entropy etc. the main propose are to determine the performance of cluster.

Table 1: Summery of Approaches Using Semantic Similarity Based On Clustering

No.	papers	Date	Method	classification and clustering Algorithm	Evaluation measures	Other Measures	Dataset
1.	L. Sahni, et al. ,[17]	2014	Semantic similarity between words using WordNet	SVM	Training the model by determining a split ratio of about 70.51%.	using 4 co- occurrence measures, probability measure and measures-based approach (LCH, WUP and Resnik)	using MEN dataset for training and 3 Benchmark datasets for testing propose
2.	Y. S. Lin et al. ,[26]	2014	Text classification and clustering	SL-KNN, ML-KNN, k-means and HAC	Accuracy and Entropy in (tf-idf, word count, Z-score)	SMTP proposed measure	Three data sets namely RCV1, WebKB and Reuters-8
3.	S. Nourashrafeddin et al. ,[29]	2014	Document clustering using Wikipedia	fuzzy c-means and Naive Bayes classifier	Normalized Mutual Information (NMI)	Var-TFIDF	using 8 datasets 1-20Newsgroups (News-sim3, News-rel3 and20ng-whole). 2-Reuters-21578. 3- from data repository SMART creating Classic4. 4-WebKaB. 5-SMS Spam Collection 6- Cade from Brazilian web pages content is gathered
4.	H. M. Alghamdi et al ,[30]	2014	Arabic VerbNet	K-Mean	Purity, Mean intra cluster distance (MICD) and Davis Bouldin Index (DBI)	using Mean score among cluster, Mean ratio among corpus (MRAC) for Annotation	Corpus collected from online Arabic newspaper archive
5.	T. Wei et al. ,[31]	2014	Text clustering Using WordNet version 2.0	Bisecting k-mean	Purity, Entropy and F1-measure	Using Lexical chains	Reuters-21578 corpus
6.	S. S. Desai and J. A. Laxminarayana ,[9]	2016	Document clustering Using WordNet	Bisecting K-mean	Purity	Tf-idf for feature selection	using 4 classic Datasets 1-60 Documents from CACM, 2- 44 Documents from CISI, 3-44 Documents from CRAN, 4- 52 Documents from MED
7.	T. Gowda and C. Mattmann ,[35]	2016	Webpage clustering using (DOM)tree	Near Neighbor Clustering technique	using different value of threshold	TED and JS	Dataset from a popular weapons classifieds site
8.	N. Fiorini et al. ,[37]	2016	Semantic clustering using WordNet	HAC	Normalized Robinson-Foulds Distance	lin Measure, BMA, post processing	Derived from WordNet-based disambiguation version
9.	L. Stanchev ,[40]	2016	Document clustering Based on similarity Graph	K-mean	F-measure	cosine, linear, and logarithmic	Reuters-21578 benchmark
10.	C. Kulathunga and D. D. Karunara,[41]	2017	Document Clustering Using WordNet and information content file	K-mean and sIB	Purity	Tf-idf	EMMA
11.	Z. Wu et al. ,[43]	2017	Document clustering Using Wikipedia	K-mean	Selectivity, relevance, purity and vector construction time	Heuristic selection	Wikipedia Dataset published from internet and document dataset from Reuters-21578
12.	J. Avanija et al. ,[44]	2017	Document Clustering using KIM plugging tool	Fuzzy C-mean with PSO and K-mean with PSO	Purity and F-Measure	Tf-idf	Web pages collect from internet
13.	S. Wang and R. Koopman ,[45]	2017	Article clustering based on Network	K-means and Louvain method	F1-measure and adjustedF1	vectoring	Astro dataset

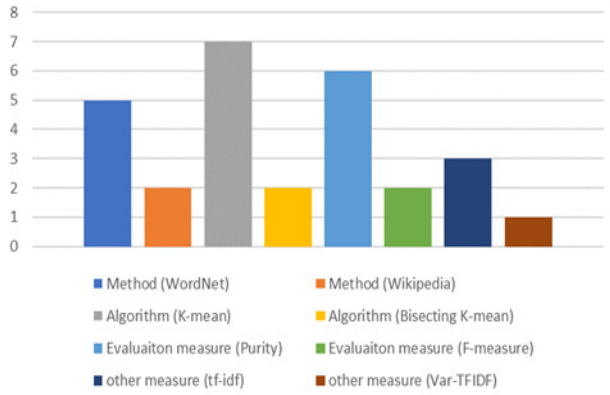


Figure 1: Showing the usage number of main features (Methods, Algorithms and measures).

4. Conclusion

In this paper, after we reviewed all that papers, we conclude that each paper has specific approaches and various tools and measures are available for each approach, and as it can be seen that most popular steps are used in approaches were preprocessing to transform the document in better format and different steps are used in preprocessing like removing stop words, stemming, tokenization etc., word sense disambiguation are used for solving the synonymy and polysemy problems and feature selection another important step that many of approaches was using tf-idf for this case and there was another like Var-TFIDF, heuristic selection etc. however one of the most popular tools used is WordNet the English dataset for meaningful clustering, then clustering done by using clustering algorithms also the most usage one was K-mean algorithm because of its simplicity to use and can give tighter cluster as well as there is another like bisecting K-mean, fuzzy c-mean, hierarchical agglomerative clustering etc. finally, as we see that the main goal of all approaches trying to a chive a better efficiency, accuracy and quality of clustering. after using the experimental measure for evaluating propose like (Purity, F-measure, Entropy etc.) in all case show that the semantic clustering giving a better performance.

References

[1] B. Everitt and Wiley InterScience (Online service), Cluster Analysis. Chichester, West Sussex, U.K: Wiley, 2011.
[2] D. Q. Zeebaree, H. Haron, A. M. Abdulazeez, and S. R. M. Zeebaree, "Combination of k-means clustering with genetic algorithm: A review," Int. J. Appl. Eng. Res., vol. 12, no. 24, pp. 14238–14245, 2017.
[3] W. L. Chang, K. M. Tay, and C. P. Lim, "A New Evolving Tree-Based Model with Local Re-learning for Document Clustering and Visualization," Neural Process. Lett., vol. 46, no. 2, pp. 379–409, 2017.
[4] K. Jacksi, N. Dimililer, and S. R. M. Zeebaree, "a Survey of Exploratory Search Systems Based on Lod Resources," Turkey. Univ. Utara Malaysia, no. 112, pp. 501–509, 2015.
[5] H. H. Tar and T. Nyunt, "Enhancing Traditional Text Documents Clustering based on Ontology," Int. J. Comput. Appl., vol. 33, no. 10, pp. 38–42, 2011.
[6] Z. Chengzhi and X. Hongjiao, "Clustering description extraction based on statistical machine learning," Proc. - 2008 2nd Int. Symp. Intell. Inf. Technol. Appl. IITA 2008, vol. 2, pp. 22–26, 2008.
[7] K. Jacksi, S. R., and N. Dimililer, "LOD Explorer: Presenting the Web of Data," Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 1, pp. 45–51, 2018.
[8] K. Jacksi, N. Dimililer, and S. R., "State of the Art Exploration Systems for Linked Data: A Review," Int. J. Adv. Comput. Sci. Appl., 2016.
[9] S. S. Desai and J. A. Laxminarayana, "WordNet and Semantic similarity-based approach for document clustering," 2016 Int. Conf. Comput. Syst. Inf. Technol. Sustain. Solut. CSITSS 2016, pp. 312–317, 2016.
[10] S. Harispe, S. Ranwez, S. Janaqi, and J. Montmain, "Semantic Similarity from Natural Language and Ontology Analysis," Synth. Lect. Hum. Lang. Technol., vol. 8, no. 1, pp. 1–254, May 2015.
[11] Y. Feng, E. Bagheri, F. Ensan, and J. Jovanovic, "The state of the art in semantic relatedness: a framework for comparison," Knowl. Eng. Rev., pp. 1–30, Mar. 2017.
[12] F. M. Couto and A. Lamurias, "Semantic Similarity Definition," in Encyclopedia of Bioinformatics and Computational Biology, Elsevier, 2019, pp. 870–876.
[13] A. Ballatore, M. Bertolotto, and D. C. Wilson, "An evaluative baseline for geo-semantic relatedness and similarity," Geoinformatica, vol. 18:4, pp. 747–767, 2014.
[14] "Semantic similarity." [Online]. Available: https://en.wikipedia.org/wiki/Semantic_similarity. [Accessed: 27-Nov-2018].
[15] A. A.-Z. K. J. A. S. Subhi R. M. Zeebaree, "Designing an Ontology of E-learning system for Duhok Polytechnic University Using Protégé OWL Tool," J. Adv. Res. Dyn. Control Syst., vol. Volume 11, no. 05-Special Issue, pp. 24–37, 2019.
[16] M. A. Hadj Taieb, M. Ben Aouicha, and A. Ben Hamadou, "Ontology-based approach for measuring semantic similarity," Eng. Appl. Artif. Intell., vol. 36, pp. 238–261, 2014.
[17] L. Sahni, A. Sehgal, S. Kochar, F. Ahmad, and T. Ahmad, "A Novel Approach to Find Semantic Similarity Measure between Words," Proc. - 2014 2nd Int. Symp. Comput. Bus. Intell. ISCBI 2014, pp. 89–92, 2015.
[18] C. Leacock and M. Chodorow, "Combining Local Context and WordNet Similarity for Word Sense Identification," WordNet An Electron. Lex. database., no. January 1998, pp. 265–283, 1998.
[19] W. Zhibiao and M. Palmer, "VERB SEMANTICS AND LEXICAL Zhibiao W u," Proc. 32nd Annu. Meet. Assoc. Comput. Linguist., pp. 133–138, 1994.
[20] P. Resnik, "Using Information Content to Evaluate Semantic Similarity in a Taxonomy."
[21] K. Sparck Jones and P. Willett, Readings in information retrieval. Morgan Kaufman, 1997.
[22] G. A. Miller and W. G. Charles, "Contextual correlates of semantic similarity," Lang. Cogn. Process., vol. 6, no. 1, pp. 1–28, Jan. 1991.
[23] H. Rubenstein and J. B. Goodenough, "Contextual correlates of synonymy," Commun. ACM, vol. 8, no. 10, pp. 627–633, Oct. 1965.
[24] L. Finkelstein et al., "Placing search in context: the concept revisited," ACM Trans. Inf. Syst., vol. 20, no. 1, pp. 116–131, 2002.
[25] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA Data Mining Software: An Update."
[26] Y. S. Lin, J. Y. Jiang, and S. J. Lee, "A similarity measure for text classification and clustering," IEEE Trans. Knowl. Data Eng., vol. 26, no. 7, pp. 1575–1590, 2014.
[27] "Datasets for single-label text categorization - Ana Cardoso Cachopo's Homepage." [Online]. Available: http://ana.cachopo.org/datasets-for-single-label-text-categorization. [Accessed: 05-Dec-2018].
[28] D. D. Lewis, Y. M. Yang, T. G. Rose, and F. Li, "RCV1: A new benchmark collection for text categorization research," J. Mach. Learn. Res., vol. 5, pp. 361–397, 2004.
[29] S. Nourashrafeddin, E. Milios, and D. V. Arnold, "An ensemble approach for text document clustering using Wikipedia concepts," Proc. 2014 ACM Symp. Doc. Eng. - DocEng '14, pp. 107–116, 2014.
[30] H. M. Alghamdi, A. Selamat, and N. S. Abdul Karim, "Arabic web pages clustering and annotation using semantic class features," J. King Saud Univ. - Comput. Inf. Sci., vol. 26, no. 4, pp. 388–397, 2014.
[31] T. Wei, Y. Lu, H. Chang, Q. Zhou, and X. Bao, "A semantic approach for text clustering using WordNet and lexical chains," Expert Syst. Appl., vol. 42, no. 4, pp. 2264–2275, 2015.
[32] H. Andreas, S. Staab, and G. Stumme, "Wordnet improves Text Document Clustering," 2003.
[33] L. Jing, M. K. Ng, and J. Z. Huang, "Knowledge-based vector space model for text clustering," Knowl. Inf. Syst., vol. 25, no. 1, pp. 35–55, Oct. 2010.
[34] S. Fodeh, B. Punch, and P.-N. Tan, "On ontology-driven document clustering using core semantic features," Knowl. Inf. Syst., vol. 28, no. 2, pp. 395–421, Aug. 2011.
[35] T. Gowda and C. Mattmann, "Clusteringweb pages based on structure and style similarity," Proc. - 2016 IEEE 17th Int. Conf. Inf. Reuse Integr. IRI 2016, pp. 175–180, 2016.
[36] R. A. Jarvis and E. A. Patrick, "Clustering Using a Similarity Measure Based on Shared Near Neighbors," IEEE Trans. Comput., vol. C–22, no. 11, pp. 1025–1034, Nov. 1973.
[37] N. Fiorini, S. Harispe, S. Ranwez, J. Montmain, and V. Ranwez, "Fast and reliable inference of semantic clusters," Knowledge-Based Syst., vol. 111, pp. 133–143, 2016.
[38] R. Bernardi and I. AT4DL 2009 (2009: Trento, Advanced language technologies for digital libraries: international workshops on NLP4DL 2009,

Viareggio, Italy, June 15, 2009 and AT4DL 2009, Trento, Italy, September 8, 2009. Springer, 2011.
[39] "SC - Semantic clustering." [Online]. Available: http://sc.nicolasfiorini.info/. [Accessed: 13-Dec-2018].
[40] L. Stanchev, "Semantic Document Clustering Using a Similarity Graph," Proc. - 2016 IEEE 10th Int. Conf. Semant. Comput. ICSC 2016, pp. 1–8, 2016.
[41] C. Kulathunga and D. D. Karunaratne, "An -ontology-based and domain specific clustering methodology for financial documents," 17th Int. Conf. Adv. ICT Emerg. Reg. ICTer 2017 - Proc., vol. 2018-Janua, pp. 209–216, 2018.
[42] "Municipal Securities Rulemaking Board:EMMA." [Online]. Available: https://emma.msrb.org/. [Accessed: 13-Dec-2018].
[43] Z. Wu et al., "An efficient Wikipedia semantic matching approach to text document classification," Inf. Sci. (Ny)., vol. 393, pp. 15–28, 2017.
[44] J. Avanija, G. Sunitha, and K. R. Madhavi, "Semantic Similarity based Web Document Clustering Using Hybrid Swarm Intelligence and FuzzyC-Means," vol. 7, no. 5, pp. 2007–2012, 2017.
[45] S. Wang and R. Koopman, "Clustering articles based on semantic similarity," Scientometrics, vol. 111, no. 2, pp. 1017–1031, 2017.
[46] J. Gläser, W. Glänzel, and A. Scharnhorst, "Same data—different results? Towards a comparative approach to the identification of thematic structures in science," Scientometrics, vol. 111, no. 2, pp. 981–998, May 2017.
[47] "WordNet Application."
[48] Santini, "Advantages & Disadvantages of k- - - Means and Hierarchical clustering (Unsupervised Learning)," pp. 1–5, 2016.
[49] A. Rajaraman and J. D. Ullman, "Data Mining," in Mining of Massive Datasets, Cambridge: Cambridge University Press, 2011, pp. 1–17.
[50] S. C. Sripada, "Comparison of Purity and Entropy of K-Means Clustering and Fuzzy C Means Clustering," Indian J. Comput. Sci. Eng., vol. 2, no. 3, pp. 343–346, 2011.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 5, 123-131 (2019)
www.astesj.com
Special Issue on Multidisciplinary Sciences and Engineering

ASTES Journal
ISSN: 2415-6698

Green Inventory Supply Chain Model with Inflation under Permissible Delay in Finite Planning Horizon

Seema Saxena[*,1], Vikramjeet Singh[2], Rajesh Kumar Gupta[1], Nitin Kumar Mishra[1], Pushpinder Singh[3]

[1]*Department of Mathematics, Lovely Professional University, Phagwara 144411, Punjab, India*

[2]*Department of Mathematics, IKGPTU Campus,Batala, 143506, Punjab, India*

[3]*Department of Mathematics, Mata Gujri College, Fatehgarh Sahib, 140407, Punjab, India*

ARTICLE INFO

Article history:
Received: 12 July, 2019
Accepted: 19 August, 2019
Online: 24 September, 2019

Keywords:
Inflation
Supply chain
Credit term
Remanufacturing

ABSTRACT

Trade credit is an important cost reduction tool in the inventory management. The effect of trade credit is studied on the integrated system for sharing the cost benefits realized due to the permissible delay. Credit term factor is introduced to divide the cost benefits between the retailer and the supplier. The various costs in the inventory model are subjected to the same inflation rate. This research paper revisits EOQ model for remanufacturing process under green supply chain with the permissible delay available to the retailer. Numerical examples prove that the optimal re-ordering schedule exists and is unique. Also sensitivity analysis is performed on certain parameters to ascertain their logical implications.

1 Introduction

This paper is an extension of the work [1] originally presented in the 4th International Conference on Computing Science(ICCS),2018. This extended research work incorporates the green inventory concept. The harmful effects of the waste and outdated products are imposing a threat to the environment. In present scenario, world is facing pollution as a big hazard to mankind. Every organization is moving towards reducing and reusing the waste/imperfect goods. In this direction this EOQ model aims at a single stage remanufacturing process where the imperfect goods after screening are taken back by the supplier and are remanufactured. This is again transported to the retailer. Replenishment schedule is derived for the retailer and the supplier, considering green product life cycle and the time value of money. Various countries have adopted several measures for waste product management, reusing and remanufacturing programs. International standards as European Union's proposal for Waste Electrical and Electronic equipment (WEEE) introduces the concept of product design, to bring a decrease in the cost of disassembly and remanufacturing [2]. Product, if designed significantly reduces the cost of inspection, disassembly, repair, remanufacturing and recycling. In [3], the author first determine an optimal ordering size with recovery and remanufacturing. He studied the traditional EOQ model with continuous and deterministic demand and return. In [4], the authors probed the changes taking place in the market and in the manufacturing organizations. He studied the changes for the process of product design, with respect to the technology introduced for new material and new production methods, including tools and techniques altered for the manufacturing, inspection, reusing and remanufacturing. In [5], the author provided a comprehensive and immense knowledge for the remanufacturing process from the literature surveyed. The supplier and retailer are stationed far from each other so it is not possible for the supplier to send all perfect goods. Thus to be assure of the brand and quality, retailer screens the lot as it is received. In [6], the authors studied the remanufacturing of the imperfect items to maximize the total profit. The effect of deterioration is dominant and its consequences cannot be ignored while framing an EOQ model. Electronic goods, blood, fruits, grain products, alcohol are some of the deteriorating products. In [7], the authors were the first to study deterioration in an inventory model. Several researchers as [8], [9], [10]. [11] studied different patterns of deterioration in inventory models. In [12], the authors developed an inventory model for the stock dependent demand for deteriorating products under two level credit. In [13], the researcher studied a two-echelon supply chain model for deteriorating products under trade credit. They analyzed two models, one with demand being stock dependent and the second with demand being selling price dependent for the perishable products. In [14], the authors studied

[*]Corresponding Author Seema Saxena, Department of Mathematics, Lovely Professional University, Phagwara 144411, Punjab(India), 8591188250 & mseemamishra@gmail.com

an inventory model for the deteriorating products where the deterioration being time dependent. They studied permissible delay policy in their model. Partially backlogged shortages were als incorporated. In [15], the authors discusses the inventory model for perishable items where deterioration follows exponential distribution under permissible delay.
In conventional EOQ model, assumes that the retailer must pay for the goods as they arrive. But as practice, sometimes the supplier allows permissible delay to the retailer for the settlement of the goods. Numbers of research papers have been published with inventory models under trade credit. The rudest approach of an inventory model under trade credit was done by [16]. [17], developed an EOQ model under two level supply chain with trade credit for deteriorating products and demand being stock dependent. In [18],[19], the authors revisited the inventory models with time and stock dependent demand for imperfect and perishable products allowing trade credit. [20] studies the effect of sales effort on demand in a supply chain inventory model when permissible delay is allowed. He also included quantity discounts in his model with two level supply chain. Trade credit is widely used by the industries in U.S., China and Europe as discussed in [21],[22] and [23]. Suppliers give credit period to the retailer to postpone their payments, thereby increasing their participation and retaining the market. On the other hand, retailer can collect the revenue of the goods during the credit period and is encouraged to increase the order quantity.The concept of trade credit incorporated with the green inventory and weibull deterioration was studied by [24]. In [25], the authors analyses an inventory system with trade credit for imperfect goods and demand being quality dependent. He has considered two different approaches of trade credit in the supply chain model.

The formal EOQ Model establishing various results does not include the inflation for all the inventory related costs, whereas its effect should not be ignored. The first initiation for an inventory model under the time value of money was considered by [26]. In [27], the researchers analyzed an inventory model with the time value of money and weibull deterioration. In [28], the authors proposed a deteriorating EOQ model when demand is poisson and lead time is non zero. In [29], the authors developed a replenishment model for perishable products with return policy under inflation. In [30], the authors determine a reordering policy in an inventory model for deteriorating goods,partial backordering under inflation. In [1], the authors represented supply chain inventory model with inflation for time quadratic demand and deterioration. In [31], the authors developed an inventory model for deteriorating products considering inflation in a green supply chain for remanufacturing and recycling. Also a fuzzified model was developed by [32].

2 Assumptions and Terminology

2.1 Assumptions

The following assumptions are followed:

1. All inventory related cost are subjected to the same constant rate of inflation r.
2. The supplier offers the retailer credit period ρ to settle the account.
3. Finite replenishment and non zero lead time is considered.
4. The available inventory deteriorates with a constant fraction θ_1 in a finite planning horizon.
5. Single item single retailer and single supplier is considered in the supply chain.
6. A percentage p of products received in the lot are of imperfect quality.
7. The model is proposed in a finite planning horizon H, having I_{oi} units at the start of the cycle.

2.2 Terminology

1. The Opportunity cost $C_c(\$/unit/year)$ is same for the retailer and the supplier.
2. M_j^S is the permissible delay given by the supplier to the retailer for the settlement of the account in the synchronized model.
3. Setup cost for the supplier is S_s(\$/order).
4. The purchasing price for the supplier and the retailer is P_s(\$/unit) and P_o(\$/unit) with $P_s < P_o$.
5. In the synchronized model, the rate at which the additional cost is distributed by both the divisions is ρ.
6. Holding cost of the retailer and the suppler is H_o and H_s respectively.
7. Deterioration and Screening cost are DC and SC respectively.
8. The total inventory carried during a cycle is R_i^D in desynchronized model.
9. Total quantity in desynchronized model is Q^D.
10. In green inventory synchronized model, TC, REM and DSM are transportation cost, re-manufacturing cost and disassembly cost respectively.
11. In desynchronized model, the present estimate of the total cost for the retailer and the supplier is $PETC_r^D$ and $PETC_s^D$.

3 Model presentation and Analysis

The model is developed for two different scenarios:
Scenario 1 : Desynchronized Model

The retailer orders the quantity as per his convenience and the supplier follows the replenishment schedule of the retailer under inflation and green's technology of remanufacturing and recycling.
Scenario 2: Synchronized Model under permissible delay

The supplier offers credit period to the retailer to pay back the money for the lot received. Also shares the extra cost incurred by the retailer on account of following the replenishment schedule of the supplier with the order size increased. The present estimate of the total cost for the retailer and the supplier is calculated under inflation and green's technology of remanufacturing and recycling.

Model Description

The lot received at the beginning of the cycle has p percentage of imperfect quality goods. The complete lot undergoes screening process by the retailer and the defective goods are taken back by the supplier for remanufacturing and recycling. The cycle starts at t_i with I_{oi} units. The fall in the inventory is due to demand and deterioration till $t_i^{'}$.The level of inventory at this first stage of the process is $I_{1i}(t)$. At $t_i^{'}$, a decrease in demand occurs when the imperfect quality goods are taken back by the supplier. From $t_i^{'}$ to $t_i^{''}$, the fall in inventory is due to demand and deterioration. $I_{2i}(t)$ is the inventory for the second stage of the process. At $t_i^{''}$, an increase in the level of inventory is observed when the imperfect goods remanufactured are transported back to the retailer. Further at $t_i^{''}$, the inventory level $I_{3i}(t)$ declines due to demand and deterioration till it reaches zero. This behaviour is illustrated by figure 1.

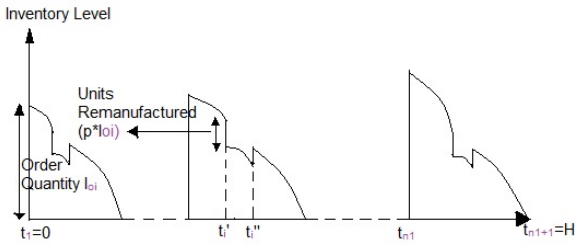


Figure 1: Graphical representation of Inventory Model

Scenario 1: Desynchronized Model

$$\begin{aligned}&\frac{d\left(I_{1i}(t)\right)}{dt}+\theta_1(t)I_{1i}(t)\\&=-f(t), t_i \le t\\&\le t_i^{'}, (i\\&=1,2,\ldots,n_1)\end{aligned} \tag{1}$$

and the boundary condition is $I_{1i}\left(t_i\right)=I_{oi}$ and $I_{1i}\left(t_i^{'}\right)=I_{si}$

$$\begin{aligned}&\frac{d\left(I_{2i}(t)\right)}{dt}+\theta_1(t)I_{2i}(t)\\&=-f(t), t_i^{'} \le t\\&\le t_i^{''}, (i\\&=1,2,\ldots,n_1)\end{aligned} \tag{2}$$

and the boundary condition is $I_{2i}\left(t_i^{'}\right)=I_{si}-p.I_{oi}$ and $I_{2i}\left(t_i^{''}\right)=I_{fi}$

$$\begin{aligned}&\frac{d\left(I_{3i}(t)\right)}{dt}+\theta_1(t)I_{3i}(t)\\&=-f(t), t_i^{''} \le t\\&\le t_{i+1}, (i\\&=1,2,\ldots,n_1)\end{aligned} \tag{3}$$

the boundary condition is $I_{3i}\left(t_i^{''}\right)=I_{fi}+p.I_{oi}$ and $I_{3i}\left(t_{i+1}\right)=0$
Solving these equations, we get

$$\begin{aligned}I_{\mathrm{oi}}=&\,a\left(t_{i+1}-t_i\right)+\left((\mathrm{a}\theta_1+b)/2\right).\left(t_{i+1}^2-t_i^2\right)+\left(\mathrm{b}\theta_1/3\right)\left(t_{i+1}^3-t_i^3\right)\\&-\left(\mathrm{p}\theta_1\left(t_{i+1}^{''}-t_i^{'}\right)+\theta_1 t_i\right).\left(a\left(t_{i+1}-t_i\right)+(b/2).\left(t_{i+1}^2-t_i^2\right)\right)\end{aligned} \tag{4}$$

$$\begin{aligned}I_{si}=&\,e^{-\theta_1.t_i^{'}}\left(1-p\left(-e^{-\theta_1\left(t_i-t_i^{'}\right)}+e^{-\theta_1\left(t_i-t_i^{''}\right)}\right)\right).\\&\int_{t_i}^{t_{i+1}}e^{\theta_1.u}.f(u)du-\int_{t_i}^{t_i^{'}}e^{\theta_1.\left(u-t_i^{'}\right)}.f(u)du\end{aligned} \tag{5}$$

$$\begin{aligned}I_{fi}=&\,I_{oi}.e^{\theta_1.\left(t_i-t_i^{''}\right)}-e^{-\theta_1 t_i^{''}}.\int_{t_i}^{t_i^{'}}e^{\theta_1.u}.f(u)du\\&-p.I_{oi}.e^{\theta_1.\left(t_i^{'}-t_i^{''}\right)}-e^{-\theta_1 t_i^{''}}.\int_{t_i^{'}}^{t_i^{''}}e^{\theta_1.u}.f(u)du\end{aligned} \tag{6}$$

substituting these values, the solution of the differential eq (1) is

$$\begin{aligned}I_{1i}(t)=&\,a\left(t_{i+1}-t\right)+\left((a.\theta_1+b)/2\right).\left(t_{i+1}^2-t^2\right)\\&+\left((b.\theta_1)/3\right).\left(t_{i+1}^3-t^3\right)+\left(p.\theta_1.t_i^{'}-p.\theta_1.t_i^{''}\right).\left(a\left(t_{i+1}\right.\right.\\&\left.\left.-t_i\right)+(b/2).\left(t_{i+1}^2-t_i^2\right)\right)-\theta_1.t.\left(a\left(t_{i+1}-t\right)\right.\\&\left.+(b/2).\left(t_{i+1}^2-t^2\right)\right), t_i \le t \le t_i^{'}, \quad \{i=1,2,\ldots,n_1\}.\end{aligned} \tag{7}$$

the solution of the differential eq (2) is $I_{2i}(t)=$
$(1-\theta_1.t).\left(a\left(t_{i+1}-t\right)+(b/2).\left(t_{i+1}^2-t^2\right)\right)+$
$\left((a.\theta_1)/2\right).\left(t_{i+1}^2-t^2\right)+\left((b.\theta_1)/3\right).\left(t_{i+1}^3-t^3\right)$
$-p.\left(\left((a.\theta_1)/2\right).\left(t_{i+1}^2-t_i^2\right)+\left((b.\theta_1)/3\right).\left(t_{i+1}^3-t_i^3\right)\right)-$

$$\begin{aligned}&p.\left(1+\theta_1.t_i^{''}-\theta_1.t-\theta_1*t_i\right).\left(a\left(t_{i+1}-t_i\right)+(b/2).\left(t_{i+1}^2-t_i^2\right)\right)\\&+p^2\left(\theta_1.t_i^{''}-\theta_1.t_i^{'}\right), t_i^{'} \le t \le t_i^{''}, \quad \{i=1,2,\ldots,n_1\}\end{aligned} \tag{8}$$

the solution of the differential eq (3) is

$$\begin{aligned}I_{3i}(t)=&\,(1-\theta_1.t).\left(a\left(t_{i+1}-t\right)+(b/2).\left(t_{i+1}^2-t^2\right)\right)+\\&\left(\left((a.\theta_1)/2\right).\left(t_{i+1}^2-t^2\right)+\left((b.\theta_1)/3\right).\left(t_{i+1}^3-t^3\right)\right), t_i^{''} \le t \le t_{i+1},\\&\{i=1,2,\ldots,n_1\}.\end{aligned} \tag{9}$$

Total inventory carried during the interval $[t_i, t_{i+1}]$ is

$$\begin{aligned}R_i^D=&\int_{t_i}^{t_i^{'}}I_{1i}(t)dt+\int_{t_i^{'}}^{t_i^{''}}I_{2i}(t)dt+\int_{t_i^{''}}^{t_{i+1}}I_{3i}(t)dt, \quad (i\\&=1,2,3,\ldots,n_1).\end{aligned} \tag{10}$$

Total quantity during the planning horizon =

$$Q^D=\sum_{i=1}^{n_1}I_{oi}^D \tag{11}$$

Retailer's total cost includes present estimate of the ordering cost, holding cost, deterioration cost, purchasing cost and screening cost. Present estimate of the retailer's total cost $PETC_r^D$ is

$$PETC_r^D=\sum_{i=1}^{i=n_1}\left(PEOC+PEHC+PEDC+PEPC+PESC\right)$$

$$\sum_{i=1}^{i=n_1} e^{(d-r)t_i}.n_1.O_r + H_o.$$
$$\int_{t_i}^{t_i'} e^{(d-r).t}.I_{1i}(t)dt + H_o.\int_{t_i'}^{t_i''} e^{(d-r).t}.I_{2i}(t)dt +$$
$$H_o.\int_{t_i''}^{t_{i+1}} e^{(d-r).t}.I_{3i}(t)dt + \text{DC}.\int_{t_i}^{t_i'} e^{(d-r).t}.\theta_1.I_{1i}(t)dt$$
$$+ \text{DC}.\int_{t_i'}^{t_i''} e^{(d-r).t}.\theta_1.I_{2i}(t)dt$$
$$+ \text{DC}.\int_{t_i''}^{t_{i+1}} e^{(d-r).t}.\theta_1.I_{3i}(t)dt + (P_o + SC).e^{(d-r).t_i}.I_{oi} \quad (12)$$

By determining the $t_i's$, the green's inventory model is optimized. Equation the first order derivative of the total cost to zero, the values are calculated.

$$\frac{\partial PETC_r^D(t_i;n_1)}{\partial t_i} = 0 \quad (13)$$

The optimal solution so obtained by solving the above equation is $n_1^{DO}, t_0, t_1^{DO}, t_2^{DO},\ldots,t_{n_1+1}^{DO} = H,$

With the existing replenishment schedule in the desynchronized model, the supplier determines his/her present estimate of the total cost by summing the setup cost,purchasing cost, holding cost, transportation cost, dissemmbly cost and the remaufacturing cost for the planning horizon H. The equation is as follows:

$$PETC_s^D =$$
$$n_1^{DO}.e^{(d-r)t_i}.S\,s + P_s.e^{(d-r)t_i}.I_{oi} + H_s.\int_{t_i'}^{t_i''} e^{(d-r)t}.p.I_{oi}\,dt +$$
$$(TC + DSM + REM).p.e^{(d-r)t}.I_{oi} \quad (14)$$

The optimized quantity to be ordered is:

$$Q^* = \sum_{i=1}^{n_1^{DO}} I_{oi}^{DO} \quad (15)$$

Scenario 2: Synchronized model with permissible delay

in the synchronized model supplier offers permissible delay to the retailer. On account of which the number of reordering cycle decreases,which further leads to a decrease in the setup cost of the supplier. With the new ordering plan, there is an increase in the cost for the retailer which the supplier compensates by distributing it through the parameter credit period rate.

For the new ordering plan n_2, the present estimate of the cost for the retailer is

$$PETC_r^S =$$
$$\sum_{j=1}^{j=n_2} (PEOC + PEHC + PEDC + PEPC + PESC) \quad (16)$$

The addition in the retailer's cost is given by:

$$PETC_r^S - PETC_r^D \quad (17)$$

The supplier compensates the retailer by adding the increase to his/her total cost and so the present estimate of the cost is

$$PETC_s^S =$$
$$n_2.e^{(d-r)t_j}.S\,s + P_s.e^{(d-r)t_j}.I_{oj} + H_s.\int_{t_j'}^{t_j''} e^{(d-r)t}.p.I_{oj}dt +$$
$$(TC + DSM + REM).p.e^{(d-r)t}.I_{oj} + PETC_r^S - PETC_r^D \quad (18)$$

The optimized value of the cost for the supplier $PETC_s^S$ in the synchronized model is determined in the same way as done for the retailer $PETC_r^D$ in the desynchronized model.

$n_2^{SO}, t_1^{SO}, t_2^{SO},\ldots,t_{n_2+1}^{SO} = H$ be the optimal solution for $PETC_s^S$.

The optimized ordering quantity in synchronized model is

$$Q^S = \sum_{j=1}^{n_2^{SO}} I_{oj}^{SO}$$

Equitable distribution of the extra cost incurred during synchronized model

For the retailer,the present estimate of the total cost in the synchronized model should be less than that in the desynchronized model, only then the new replenishment program is followed by the retailer.

$$PETC_r^D \geq PETC_r^S - \sum_{j=1}^{n_2^{SO}} e^{(d-r)*t_j}C_c.\rho\left(t_{j+1}^{SO} - t_j^{SO}\right)I_{oj}^{SO} \quad (19)$$

where ρ is the credit period rate which is the distributing factor for the extra cost at the retailer's end and the rate is same in all ordering intervals.
The credit period duration is given by:
$M_j^S = \rho\left(t_{j+1}^{SO} - t_j^{SO}\right)$
In the synchronized model,the supplier offers retailer credit for the minimum period.The credit period rate is kept minimum ρ_{min} to divide the benefits realized.The present estimate of his total cost in the synchronized model is then given by:

$$PETC_r^S - \sum_{j=1}^{n_2^{SO}} e^{(d-r)*t_j}C_c.\rho_{min}\left(t_{j+1}^{SO} - t_j^{SO}\right)I_{oj}^{SO} = PETC_r^D \quad (20)$$

which implies the value of ρ_{min} is:

$$\rho_{\min} = \frac{PETC_r^S - PETC_r^D}{\sum_{j=1}^{n_2^{SO}}.e^{(d-r)*t_j}.C_c\left(t_{j+1}^{SO} - t_j^{SO}\right)I_{oj}^{SO}} \quad (21)$$

As supplier bears the additional cost borne at the retailer's end, the credit period rate is maximum at his end ρ_{max}. His total cost in the synchronized model is given by the following equation:

$$n_2^{SO} S_s . e^{(d-r)*t_j} + \sum_{j=1}^{n_2^{SO}} P_s . e^{(d-r)*t_j} . I_{oj}^{SO} + \sum_{j=1}^{n_2^{SO}} . e^{(d-r)*t_j} . C_c .$$
$$\rho_{\max} \left(t_{j+1}^{SO} - t_j^{SO} \right) I_{oj}^{SO} = PETC_s^D \quad (22)$$

and ρ_{max} equals

$$\rho_{\max} = \frac{PETC_s^D - n_2^{SO} . e^{(d-r)*t_j} . S_s + \sum_{j=1}^{n_2^{SO}} P_s . e^{(d-r)*t_j} . I_{oj}^{SO}}{\sum_{j=1}^{n_2^{SO}} . e^{(d-r)*t_j} . C_c \left(t_{j+1}^{SO} - t_j^{SO} \right) I_{oj}^{SO}} \quad (23)$$

The additional expenditure due to the replenishment plan of the synchronized model is fractioned on the basis of $\bar{\rho}$, which is the average of the minimum and the maximum credit period rate. The present estimate of the net cost for both the retailer and the supplier is given by:

$$PETC_r^{SO\rho} = PETC_r^S \sum_{j=1}^{n_2^{SO}} . e^{(d-r)*t_j} . C_c \bar{\rho} \left(t_{j+1}^{SO} - t_j^{SO} \right) I_{oj}^{SO} \quad (24)$$

and $PETC_s^{SO\rho} =$

$$n_2^{SO} . e^{(d-r)*t_j} . S_s + \sum_{j=1}^{n_2^{SO}} P_s . e^{(d-r)*t_j} . I_{oj}^{SO} + \sum_{j=1}^{n_2^{SO}} . e^{(d-r)*t_j} . C_c . \bar{\rho} \left(t_{j+1}^{SO} - t_j^{SO} \right) I_{oj}^{SO} \quad (25)$$

4 Optimality Evaluation and Solution

The model is aimed at minimizing the total cost for the above cited two scenarios. The cycle time at which the cost is minimum is determined and the following theorem is verified with the help of the graphs.

Theorem 1: For any n_1,solution for the green supply chain EOQ model exists and is unique.

The convexity of the total cost can be seen through the following figures. Fig.2 plots the costs of retailer for the parameter SC = 4.5, and the value is minimum at the fourth cycle. The same is seen in Fig.3 for the supplier in the synchronized model where third cycle has a minimum value and the convexity of the cost is validated. The table values for the following figures is seen in Table 1 and Table 2 of the numerical example.

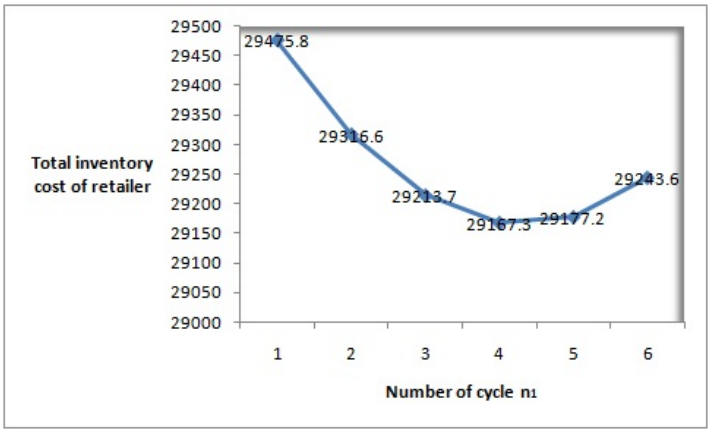


Figure 2: Convexity of total cost for retailer at SC=4.5 in desynchronized system

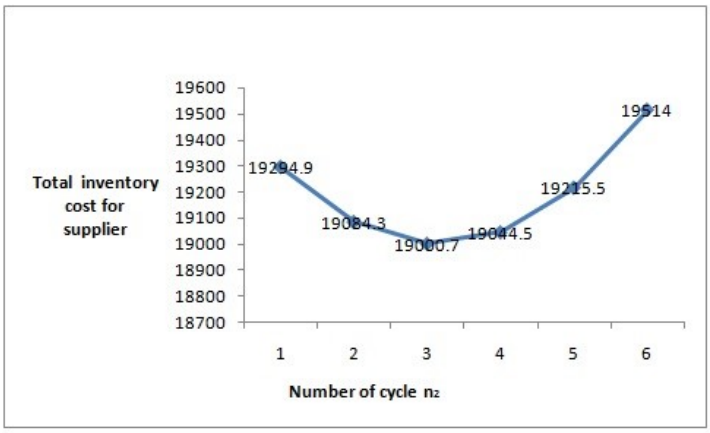


Figure 3: Convexity of total cost for supplier at SC=4.5 in synchronized system

5 Algorithm

1. All the parameters are assigned the hypothetical values.
2. In the desynchronized structure, the optimal reordering point for the retailer is determined.
   (a) Taking $t_1 = 0$, $n_1 = 1$ and $t_3 = H$. Initializing t_2, and calculating it from equation (13).
   (b) Taking $n_1 = 2$.
   (c) From the values of t_1 and t_2 the value of t_3 is calculated from equation (13).
   (d) The optimal values of $t_i's$ is determined for every n_1.
   (e) For $n_1 = 1$ and if $PETC_r^D(n_1) < PETC_r^D(n_1 + 1)$, then $PETC_r^D(n_1) = PETC_r^{DO}(n_1)$. Stop.
3. For $n_1 \geq 2$ and if $PETC_r^D(n_1) < PETC_r^D(n_1 - 1)$ and $PETC_r^D(n_1) < PETC_r^D(n_1 + 1)$, then $PETC_r^D(n_1) = PETC_r^{DO}(n_1)$ and stop or else let $n_1 = n_1 + 1$, and goto step 2(c).
4. $n_1^{DO} = n_1$ is the optimised cycle for both the retailer and the supplier in the desynchronized model.
5. $PETC_r^{DO}$, $PETC_s^{DO}$ and Q^{DO} are determined from the respective equations.
6. In synchronized model,the optimal cycle time n_2^{SO}, the present estimate of the cost for the retailer and the supplier $PETC_r^{SO}$ and $PETC_s^{SO}$ are calculated as done in steps from 2 to 4.
7. The values of $\rho_{\min}$, $\rho_{\max}$,$PETC_r^{SO\rho}$ and $PETC_s^{SO\rho}$ are determined from the respective equations.

5.1 Numerical Example

Hypothetical statistics is taken to validate the potency of the green inventory model.One principal parameter Screening Cost,SC is undertaken to observe the effects of changes in it on the optimal outcomes.

Example 1: a_1=20 items/year, b_1=5 items/year, $\theta_1 = 0.05$ items/year, $P_o = 4\$/item, \gamma = 0.01, t_1 = 0, O_r = 25$ \$/purchase order, $S_s = 30$ \$/arrangement, $H = 4$, $P_s = 0.01$ \$/item, $C_c = 2.5$ \$/item/year , $H_o = 100$ \$/item/year, d = 0.1, r = 0.05, DC = 70

$/item, CT = 0.1, TC=0.01 $/item, REM=0.01 $/item, DSM=0.01 $/item.

Table 1: Inflated cost of retailer in the desynchronized model under green's inventory system

SC	n_1		
	1	2	3
3	28568.6	28466.4	**28422.2**
3.75	28867.3	**28821**	28831.7
4.5	**29167.3**	29177.2	29243.6
SC	n_1		
	4	5	6
3	28568.6	28466.4	**28422.2**
3.75	28867.3	**28821**	28831.7
4.5	**29167.3**	29177.2	29243.6

The parameter of screening cost is further analyzed to study the impact of its changes on the green inventory model. Table 1 and Table 2 gives the total cost and the optimal ordering time period for both the retailer and the supplier in the desynchronized and the synchronized model respectively. Figure 2 and Figure 3 show the convexity of the cost.

Table 2: Inflated cost of supplier in the synchronized model under green's inventory system

SC	n_2		
	1	2	3
3	19707.5	19299.4	19020.9
3.75	19474.8	19165.2	18984
4.5	19294.9	19084.3	**19000.7**
SC	n_2		
	4	5	6
3	18872.2	**18853.3**	18964.4
3.75	**18931.1**	19006.8	19211
4.5	19044.5	19215.5	19514

Table 3 shows that the optimal number of replenishment schedules after the synchronization diminishes when the screening cost increases. This signifies that the supplier will be benefited is he/she reduces the number of ordering cycles. To add on, the percentage in the cost savings of the supplier also increases when the parameter SC increases, which is agreeable as the supplier compensates for the increase in the cost of the retailer in the synchronized system. Moreover same trend in the cost savings is seen for the retailer also. the percentage cost savings increases for the retailer when SC increases, which will assure the retailer's participation in the synchronized scheme.
Furthermore, the data also show that allowing for credit period, the optimal order quantity in the synchronized model is more than that as compared to the desynchronized model,thus raising the quantity ordered with reduced cycles.

Table 3: Percentage savings in cost for the retailer and the supplier in green's inventory system

Desynchronized system

SC	$PETC_r^{DO}$	$PETC_s^{DO}$	n_1^{DO}	Q^{DO}	$\bar{\rho}$
3	28422.2	1914.67	6	118.912	0.728528
3.75	28821	1518.23	5	118.078	0.479235
4.5	29167.3	1195.08	4	117.79	0.273768

Synchronized system

SC	$PETC_r^{SO\rho}$	$PETC_s^{SO\rho}$	n_2^{SO}	Q^{SO}
3	27730.2	1222.63	5	119.485
3.75	28131.3	828.518	4	119.485
4.5	28469.5	497.275	3	119.485

% Cost saving

SC	$\frac{\Delta PETC}{PETC_r^{DO}}$	$\frac{\Delta PETC}{PETC_s^{DO}}$
3	2.43485	36.144
3.75	2.3931	45.4288
4.5	2.39243	58.3899

6 Theoretical aspects in green inventory model on trade credit

Hypothesis 1: Permissible delay is negatively associated with the inflation rate.

To deal with the changes in the financial costs that incur due to the changing rate of inflation, organisations frequently alter the policy of trade credit. There is a decrease in the credit period to cope with the inflation. This study has been made by several researchers.The following table shows that as the rate of inflation increases, there is a decrease in the credit period rate. The data also reveals that although inflation is anticipated yet there is a realization of cost savings for both the retailer and the supplier in the green inventory model.

Table 4: Effect of change in rate of inflation on credit period rate

r	$\bar{\rho}$	$\frac{\Delta PETC}{PETC_r^{DO}}$	$\frac{\Delta PETC}{PETC_s^{DO}}$
0.025	0.807762	2.30863	33.2966
0.05	0.728549	2.43486	36.1441
0.075	0.653612	2.55951	39.0887

Hypothesis 2: Permissible delay is negatively associated to the capital cost.

Supplier incurs extra cost due to permissible delay. However, both the retailer and the supplier at distinct rates can make financial investments. So with the increase in the capital cost, there is

a reduction in the credit period rate. Furthermore,the percentage savings in the cost for the retailer and the supplier is not associated with the alterations in the capital cost.

Table 5: Effect of change in rate of Capital cost on credit period rate

Cc	$\bar{\rho}$	$\frac{\Delta PETC}{PETC_r^{DO}}$	$\frac{\Delta PETC}{PETC_s^{DO}}$
1.5	1.45327	2.4252	36.1042
3	0.728548	2.43485	36.144
4.5	0.485699	2.43485	36.144

Hypothesis 3: Credit period rate is negatively associated with the deterioration.

When the inventory of finished goods is perishable, there is a decrease in the revenue collected due to deterioration. For the deteriorated goods the time duration for permissible delay also decreases.The hypothesis is validated by the numerical example and the following table. To add in the validation for the hypothesis [33],[34] and [35] too derived the same result through the mathematical model in their research work.

Table 6: Effect of change in rate of deterioration on credit period rate

θ_1	$\bar{\rho}$	$\frac{\Delta PETC}{PETC_r^{DO}}$	$\frac{\Delta PETC}{PETC_s^{DO}}$
0.0025	0.762206	2.26233	33.3653
0.005	0.728544	2.43483	36.144
0.0075	0.696866	2.70232	40.493

Hypothesis 4: Permissible delay is positively associated with the supplier's set up cost.

There is an increase in the credit period rate when the parameter Ss increases. This guarantees that the supplier prolongs the permissible delay period to attain the benefit from synchronization. A majority of works by [36], [37], [33] predicted through their work, the attribution of the firm's profitability with respect to the increase in the delayed time and the set up cost. There is percentage cost savings fruition for the supplier which enables him to extend credit period incentive to the retailer. The retailer too generates percentage savings in the cost,thereby motivational for him to accept the synchronization scheme.

Table 7: Effect of change in supplier's set up cost on credit period rate

Ss	$\bar{\rho}$	$\frac{\Delta PETC}{PETC_r^{DO}}$	$\frac{\Delta PETC}{PETC_s^{DO}}$
15	0.536511	1.38002	30.8952
30	0.728547	2.43485	36.144
45	0.931589	3.54378	39.4326

Hypothesis 5: The change in the optimal total cost of the retailer and the supplier is negatively associated with the parameter p, but permissible delay is positively associated with p.

With the high value of the inspection rate,the imperfect goods are quickly removed,thus reducing the cost of holding. The percentage savings is negatively associated as the goods removed decreases the time period of the revenue generation. The credit period increases as it takes time for the imperfect goods to be remanufactured and get again absorbed as demand in the greens inventory model, thus validating the hypothesis.

Table 8: Effect of change in the parameter p on credit period rate

p	$\bar{\rho}$	$\frac{\Delta PETC}{PETC_r^{DO}}$	$\frac{\Delta PETC}{PETC_s^{DO}}$
0.375	0.694987	2.78661	41.9272
0.75	0.737651	2.64678	39.2901
1.125	0.778312	2.49702	36.6757

6.1 Managerial Insights

The logical implications for the hypothesis derived are firm and clear through the numerical considered and also through the work done by various prominent researchers. First, the association of the delayed cash time period with the rate of inflation is shown. This suggest that, the longer delay time should be squeezed to short time period on account of increase in the rate of inflation. Although to reach to long term conclusions, further study on other parameters associated should be done. This hypothesis highlights the significance of the supply chain process in optimising the inventory and the inventory related cost. Next in the supply chain process, the parameters as the capital cost, deterioration, set-up cost and price are studied with the delayed time. Thus modelling a concept of payment time delay with the manangement of inventory can help the firms to grab the untapped gains.

6.2 Senstivity analysis

Sensitivity analysis is carried on the green inventory model analyzing whether the formulated model is influenced by the alterations in the input parameters.We analyze the consequence of the alterations in the factors against the changes in the total cost of the retailer, supplier and the credit period rate. Every parameter is altered by -50%, -25%, 25% and 50%, of the initial cost taken in Example 1.

The table 9 shows that the change in the cost of the retailer and the supplier increases as demand increases which is agreeable. Also there is a decrease in the percentage change in the credit period with the increase in the demand.

Table 9: Analysis on demand variable 'a'

value	$\frac{\Delta PETC_r^{SO\rho} * 100\%}{PETC_{rO}^{SO\rho}}$	$\frac{\Delta PETC_s^{SO\rho} * 100\%}{PETC_{sO}^{SO\rho}}$	$\frac{\Delta\bar{\rho} * 100\%}{\bar{\rho}_O}$
0.5	-30.0024	-29.83	2.8461
0.75	-15.36	-6.576	15.2355
1.25	15.3233	5.503	-8.733
1.5	30.6353	10.7837	-15.479

The analysis on the holding cost of the supplier indicates that with the increase in the cost value the change in the cost of the supplier increases and also the rate of credit period increases.This implies that the supplier's total cost is sensitive to the holding cost as shown in Table 10.

Table 10: Analysis on holding cost of supplier H_s

value	$\frac{\Delta PETC_r^{SO\rho} * 100\%}{PETC_{rO}^{SO\rho}}$	$\frac{\Delta PETC_s^{SO\rho} * 100\%}{PETC_{sO}^{SO\rho}}$	$\frac{\Delta\overline{\rho} * 100\%}{\overline{\rho}_O}$
35	0.5753	-13.003	-15.219
52.5	0.2850	-6.449	-7.548
87.5	-0.2848	6.452	7.551
105	-0.5697	12.905	15.105

With the analysis done on the ordering cost of the retailer as in Table 11, it is evident that when the ordering size increases the total cost of the retailer increases. But there is a decrease in the cost of the supplier as the number of cycles decreases which is the synchronized model.

Table 11: Analysis on the Ordering cost O_r

value	$\frac{\Delta PETC_r^{SO\rho} * 100\%}{PETC_{rO}^{SO\rho}}$	$\frac{\Delta PETC_s^{SO\rho} * 100\%}{PETC_{sO}^{SO\rho}}$	$\frac{\Delta\overline{\rho} * 100\%}{\overline{\rho}_O}$
12.5	-0.8777	17.69	19.444
18.75	-0.444	8.96	9.851
31.25	1.110	-24.81	-27.772
37.5	1.849	-47.07	-48.271

When the parameters P_s and REM are examined,the data is recorded in Table 12 and Table 13. It is seen that with the increase in the purchasing cost and the remanufacturing cost, percentage change in the the total cost of the supplier also increases which is apparent. The supplier's total cost is sensitive to these cost parameters. The credit period rate also increases wth the increase in these parameters. Although the retailer's cost is less responsive to these cost parameters. With the increase in the credit period the retailer can invests the amount to generate gains.

Table 12: Analysis on the Purchasing cost of the supplier P_s

value	$\frac{\Delta PETC_r^{SO\rho} * 100\%}{PETC_{rO}^{SO\rho}}$	$\frac{\Delta PETC_s^{SO\rho} * 100\%}{PETC_{sO}^{SO\rho}}$	$\frac{\Delta\overline{\rho} * 100\%}{\overline{\rho}_O}$
0.005	0.0093	-0.3012	-0.2619
0.0075	-0.000293	-0.0328	-0.00151
0.0125	0.000299	0.0327	0.00136
0.015	0.000594	0.0655	0.00283

Table 13: Analysis on the Remanufacturing cost of the supplier REM

value	$\frac{\Delta PETC_r^{SO\rho} * 100\%}{PETC_{rO}^{SO\rho}}$	$\frac{\Delta PETC_s^{SO\rho} * 100\%}{PETC_{sO}^{SO\rho}}$	$\frac{\Delta\overline{\rho} * 100\%}{\overline{\rho}_O}$
0.005	0.0102	-0.2518	-0.2946
0.0075	0.0002	-0.0138	-0.0162
0.0125	-0.0001	0.0137	0.0160
0.015	-0.0003	0.0276	0.0322

7 Conclusion

The research demonstrates a synchronized supply chain with permissible delay under inflation in a greens inventory system. The model shows that the organizations can adopt for the green technology favouring the environment and can still realize cost savings. The study shows that the cost under inflation in the synchronized model is not more than the cost in the desynchronized model. The primary characteristic of the synchronized model is the credit period rate. It divides the extra cost incurred to the retailer on account of accepting the new ordering plan by the supplier and the extra cost incurred to the supplier on account of green's technology of remanufacturing and recycling. The analysis of the model reveals that both the retailer and the supplier are able to generate financial benifits. Logical insights are proposed for the various parameters in this study.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment: The authors sincerely thank and appreciate the editor-in-chief and the anonymous reviewers who have considerably helped with their observations and insightful suggestions for the refinement of the paper.

References

[1] Vikramjeet Singh, Seema Saxena, Rajesh Kumar Gupta, Nitin Kumar Mishra, and Pushpinder Singh. A supply chain model with deteriorating items under inflation. In *2018 4th International Conference on Computing Sciences (ICCS)*, pages 119–125. IEEE, 2018.

[2] Michael W Toffel. End-of-life product recovery: Drivers, prior research, and future directions. In *Conference on European Electronics Take-back Legislation: Impacts on Business Strategy and Global Trade*. Citeseer, 2002.

[3] David A Schrady. A deterministic inventory model for reparable items. *Naval Research Logistics Quarterly*, 14(3):391–398, 1967.

[4] Maurice Bonney, Svetan Ratchev, and Idir Moualek. The changing relationship between production and inventory examined in a concurrent engineering context. *International Journal of Production Economics*, 81:243–254, 2003.

[5] Marisa MP de Brito. *Managing reverse logistics or reversing logistics management?* Number ERIM PhD Series; EPS-2004-035-LIS. 2004.

[6] Biswajit Sarkar, Waqas Ahmed, Seok-Beom Choi, and Muhammad Tayyab. Sustainable inventory management for environmental impact through partial backordering and multi-trade-credit-period. *Sustainability*, 10(12):4761, 2018.

[7] P. M. Ghare and G. F. Schrader. (1963) A Model for an Exponential Decaying Inventory. Journal of Industrial Engineering, 14, 238-243. *J. Ind. Eng.*, 14: 238–243, 1963.

[8] Richard P Covert and George C Philip. An eoq model for items with weibull distribution deterioration. *AIIE transactions*, 5(4):323–326, 1973.

[9] Fred Raafat. Survey of literature on continuously deteriorating inventory models. *Journal of the Operational Research society*, 42(1):27–37, 1991.

[10] Hui-Ming Wee. Economic production lot size model for deteriorating items with partial back-ordering. *Computers & Industrial Engineering*, 24(3):449–458, 1993.

[11] H-M Wee and C-J Chung c. Optimising replenishment policy for an integrated production inventory deteriorating model considering green component-value design and remanufacturing. *International Journal of Production Research*, 47 (5):1343–1368, 2009.

[12] Kun-Jen Chung. An algorithm for an inventory model with inventory-level-dependent demand rate. *Computers & Operations Research*, 30(9):1311–1317, 2003.

[13] M Rameswari and R Uthayakumar. An integrated inventory model for deteriorating items with price-dependent demand under two-level trade credit policy. *International Journal of Systems Science: Operations & Logistics*, 5 (3):253–267, 2018.

[14] Pushpinder Singh, Nitin Kumar Mishra, Vikramjeet Singh, and Seema Saxena. An eoq model of time quadratic and inventory dependent demand for deteriorated items with partially backlogged shortages under trade credit. In *AIP Conference Proceedings*, volume 1860, page 020037. AIP Publishing, 2017.

[15] Karuppuchamy Annadurai. Integrated inventory model for deteriorating items with price-dependent demand under quantity-dependent trade credit. *International Journal of Manufacturing Engineering*, 2013, 2013.

[16] Suresh Kumar Goyal. Economic order quantity under conditions of permissible delay in payments. *Journal of the operational research society*, 36(4):335–338, 1985.

[17] Kun-Jen Chung and Leopoldo Eduardo Cárdenas-Barrón. The simplified solution procedure for deteriorating items under stock-dependent demand and two-level trade credit in the supply chain management. *Applied Mathematical Modelling*, 37(7):4653–4660, 2013.

[18] Biswajit Sarkar. An eoq model with delay in payments and time varying deterioration rate. *Mathematical and Computer Modelling*, 55(3-4):367–377, 2012.

[19] T Sarkar, SK Ghosh, and KS Chaudhuri. An optimal inventory replenishment policy for a deteriorating item with time-quadratic demand and time-dependent partial backlogging with shortages in all cycles. *Applied Mathematics and Computation*, 218(18):9147–9155, 2012.

[20] Zhihong Wang and Shaofeng Liu. Supply chain coordination under trade credit and quantity discount with sales effort effects. *Mathematical Problems in Engineering*, 2018, 2018.

[21] Tsung-Te Lin and Jian-Hsin Chou. Trade credit and bank loan: Evidence from chinese firms. *International Review of Economics & Finance*, 36:17–29, 2015.

[22] Yong-Wu Zhou, Zong-Liang Wen, and Xiaoli Wu. A single-period inventory and payment model with partial trade credit. *Computers & Industrial Engineering*, 90:132–145, 2015.

[23] Brahima Coulibaly, Horacio Sapriza, and Andrei Zlate. Financial frictions, trade credit, and the 2008–09 global financial crisis. *International Review of Economics & Finance*, 26:25–38, 2013.

[24] Vikramjeet Singh, Nitin Kumar Mishra, Sanjay Mishra, Pushpinder Singh, and Seema Saxena. A green supply chain model for time quadratic inventory dependent demand and partially backlogging with weibull deterioration under the finite horizon. In *AIP Conference Proceedings*, volume 2080, page 060002. AIP Publishing, 2019.

[25] Brojeswar Pal. Optimal production model with quality sensitive market demand, partial backlogging and permissible delay in payment. *RAIRO-Operations Research*, 52(2):499–512, 2018.

[26] JA Buzacott. Economic order quantities with inflation. *Operational research quarterly*, pages 553–558, 1975.

[27] Sh-Tyan Lo, Hui-Ming Wee, and Wen-Chang Huang. An integrated production-inventory model with imperfect production processes and weibull distribution deterioration under inflation. *International Journal of Production Economics*, 106(1):248–260, 2007.

[28] Mohammadmahdi Alizadeh, Hamidreza Eskandari, S Mehdi Sajadifar, and Christopher D Geiger. Analyzing a stochastic inventory system for deteriorating items with stochastic lead time using simulation modeling. In *Proceedings of the Winter Simulation Conference*, pages 1650–1662. Winter Simulation Conference, 2011.

[29] Maryam Ghoreishi, Gerhard-Wilhelm Weber, and Abolfazl Mirzazadeh. An inventory model for non-instantaneous deteriorating items with partial backlogging, permissible delay in payments, inflation-and selling price-dependent demand and customer returns. *Annals of Operations Research*, 226(1):221–238, 2015.

[30] Vikramjeet Singh, Seema Saxena, Pushpinder Singh, and Nitin Kumar Mishra. Replenishment policy for an inventory model under inflation. In *AIP Conference Proceedings*, volume 1860, page 020035. AIP Publishing, 2017.

[31] Smita Rani, Rashid Ali, and Anchal Agarwal. Green supply chain inventory model for deteriorating items with variable demand under inflation. *International Journal of Business Forecasting and Marketing Intelligence*, 3(1):50–77, 2017.

[32] Pushpinder Singh, Nitin Kumar Mishra, Manoj Kumar, Seema Saxena, and Vikramjeet Singh. Risk analysis of flood disaster based on similarity measures in picture fuzzy environment. *Afrika Matematika*, 29(7-8):1019–1038, 2018.

[33] Daniel Seifert, Ralf W Seifert, and Olov HD Isaksson. A test of inventory models with permissible delay in payment. *International Journal of Production Research*, 55(4):1117–1128, 2017.

[34] Abubakar Musa and Babangida Sani. Inventory ordering policies of delayed deteriorating items under permissible delay in payments. *International Journal of Production Economics*, 136(1):75–83, 2012.

[35] Xiping Song and Xiaoqiang Cai. On optimal payment time for a retailer under permitted delay of payment by the wholesaler. *International Journal of Production Economics*, 103(1):246–251, 2006.

[36] YW Zhou and D Zhou. Determination of the optimal trade credit policy: a supplier-stackelberg model. *Journal of the Operational Research Society*, 64 (7):1030–1048, 2013.

[37] Jianwen Luo and Qinhong Zhang. Trade credit: A new mechanism to coordinate supply chain. *Operations Research Letters*, 40(5):378–384, 2012.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 132-138 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Optimization Method of Wideband Multilayer Meander-Line Polarizer using Semi-Analytical approach and Application to 6-18GHz Polarizer including test with Horn Antenna

Wafa Abdouni-Abdallah*, Muhammad Saeed Khan[1], Athanasios Konstantinidis[1], Anne-Claude Tarot[2], Aziz Ouacha[1]

[1]*Emirates Technology and Innovation Center, Communication and Electronics Department, Khalifa University, Abu Dhabi, UAE*

[2]*Institute of Electronics and Telecommunications of Rennes, Antennas & Microwave Devices Department, University of Rennes 1, 35700 Rennes, France*

ARTICLE INFO

Article history:
Received: 30 May, 2019
Accepted: 28 August, 2019
Online: 24 September, 2019

Keywords:
Meander-Line Polarizer
Wideband Polarizer
Multilayer Structure
Transmission-Line Model
Simulated Annealing Algorithm

ABSTRACT

An optimization method using a semi-analytical approach is detailed in this paper to design a wideband multilayer meander-line polarizer. It consists of the transmission line equivalent circuit study to find the best values of the shunt components that allow the maximum transmission and a 90 phase difference between two orthogonal components over the entire bandwidth, followed by a simulated annealing algorithm to optimize the polarizer's dimensions by converging the values of the shunt components found earlier, to some rigorous empirical formulas of the equivalent admittances.
To validate the proposed method, a wideband (6-18 GHz) four layer's meander-line polarizer is optimized and manufactured. The polarizer prototype is stacked to a horn antenna operating at the same frequency band and tested. The experimental results verify that the linear waves of the horn antenna alone were converted to circular ones when the multilayer meander-line polarizer is stacked to it.

1. Introduction

This paper is an extension of work originally presented in the 48th European Microwave Week [1] where a four layer's wideband (6 – 18 GHz) meander-line polarizer has been optimized and simulated with different patch antenna for validation.

As well known, the polarizer is a passive broadband device, used to change the polarization properties of an Electro-Magnetic (EM) wave. Combining antennas with those devices to change the polarization of the wave is very useful. A well-designed polarizer permits to change the antenna polarization and to propagate a different polarization without any significant variations in the pattern performance [2]. The Meander-Line (ML) polarizer is the most popular device used to change the Linear Polarized (LP) waves into the Circular Polarized (CP) ones [3]. It is a multilayer structure made of several printed ML sheets spaced by about one-quarter wavelength apart. The principle of operation is to convert an incident wave into two equal components at ±45º as summarized in [4]. One component passes through a structure equivalent to a broad-band shunt-inductive filter while the other passes through a broad-band shunt-capacitive filter. the other component, by about 45º.

The ML polarizer has been proposed in [4], and improved in [5]. Later, [6-9] gave analytical formulas to compute the phase delay and characterize the polarizer's grating layer. Transmission Line (TL) theory together with Method of Moments (MOM) based algorithms has been used by [7, 10-13] to analyze the ML polarizer.

In this paper, the basic principle of the ML polarizer is described in section 2. Section 3 describes the detail of the optimization procedure which includes the TL equivalent circuit and Simulated Annealing (SA) algorithm. Finally, in section 4, the four layer's wideband (6 – 18 GHz) meander-line polarizer studied previously in [1] is used as a case study. The optimization study is presented, followed by experimental validation using a wideband horn antenna covering the same bandwidth.

2. Meander Line Polarizer: Basic Principle

The ML polarizer is a multilayer structure (Figure 1). Each layer consists of MLs printed on a low-loss dielectric substrate and separated by a spacer respecting one-quarter wavelength apart. The

*Wafa Abdouni-Abdallah, Emirates Technology and Innovation Center, Khalifa University, Main Campus, Abu Dhabi, UAE, +971 690 3881, wafa.abdouni@etic.ac.ae

spacer could be air (vacuum) or any kind of low loss material such as foam or dielectric substrate.

The principle of operation is to convert an incident wave into two equal and perpendicular components at ±45 degrees as summarized in [3]. The ML polarizer's layer is placed in the *xy*-plane in such a way that the meander axes are all oriented at an angle Ψ = 45 degrees with respect to the incident electric field $\boldsymbol{E_{inc}}$ along the *y*-axis (Figure 2).

The incident electric field $\boldsymbol{E_{inc}}$ will be decomposed into two orthogonal components: one parallel and the other one perpendicular to the ML axis, noted respectively $\boldsymbol{E_{\parallel}}$ and $\boldsymbol{E_{\perp}}$. As well known, a CP wave can be obtained by designing a ML polarizer that introduces a differential phase shift of 90^0 between the components $\boldsymbol{E_{\parallel}}$ and $\boldsymbol{E_{\perp}}$ while keeping their magnitudes identical over the entire required frequency bandwidth. Right-hand or left-hand circular polarizations can be easily obtained by switching the orientation angle $\boldsymbol{\Psi}$ from $+45^0$ to -45^0.

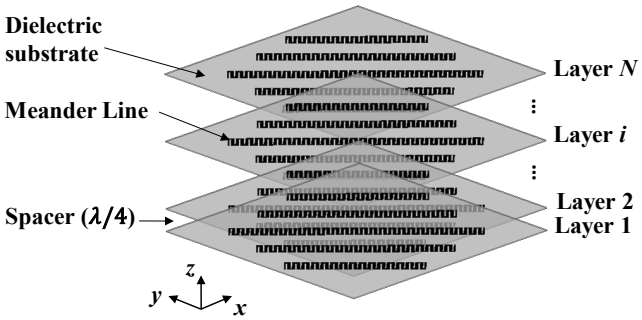


Figure 1: N-Layers ML polarizer design

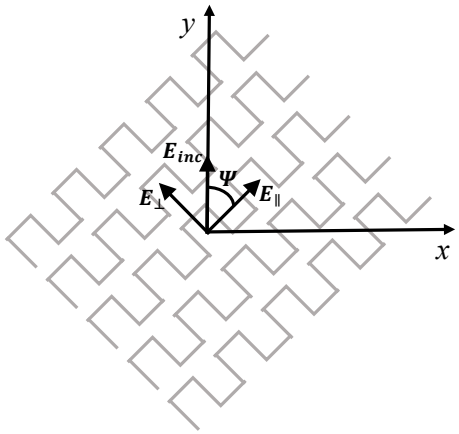


Figure 2: Incident field ($\boldsymbol{E_{inc}}$) exiting the polarizer layers with the parallel $\boldsymbol{E_{\parallel}}$and the perpendicular $\boldsymbol{E_{\perp}}$ components resulted.

The difficulty in the design of a meander-line polarizer is to determine the inductive and capacitive impedances of the grating for the $\boldsymbol{E_{\parallel}}$ and $\boldsymbol{E_{\perp}}$ polarizations with respect to frequency and constitutive dimensions of the meander line.

An optimization method is developed to optimize the ML polarizer's dimensions and described in detail in the below section. It consists of a semi-analytical analysis based on the combination of TL circuit theory and SA algorithm and allows to find the dimensions of the MLs that keep the magnitude of the components $\boldsymbol{E_{\parallel}}$and $\boldsymbol{E_{\perp}}$ identical with a phase differential shift of 90º over the desired frequency bandwidth.

Our method of optimization has the advantage of avoiding the heavy computations and simulations used in [13] to solve the TL circuit and the unit cell of each layer. This implies a large reduction in the computation time.

- The TL equivalent circuit is solved using the optimization tool of ADS. This tool has the advantage of optimizing the shunt elements of all the layers at one simulation and avoid the heavy computation of the transmission matrices of each layer then the scattering transmission parameters and finally optimize the shunt elements using quasi-Newton algorithm as used in [13].
- SA algorithm is used to optimize the physical dimensions of the MLs. This algorithm optimizes the dimensions and minimize the absolute values of the differences between the theoretical admittances introduced by Chu and Lee in [14], and their lumped-element counterparts optimized by solving the TL circuit using the optimization tool of ADS. This method avoids the full-wave analysis of the unit cell and all the steps described before which are used in [13].

3. Meander Line Polarizer: Optimization and validation Procedures

3.1. TL Equivalent Circuit Model

As summarized in [4], the incident electric field $\boldsymbol{E_{inc}}$ will be decomposed into two orthogonal components $\boldsymbol{E_{\parallel}}$ and $\boldsymbol{E_{\perp}}$. One component passes through a structure equivalent to a broad-band shunt-inductive filter while the other passes through a broad-band shunt-capacitive filter. The phase shift through either filter should have almost the same slope, so that if the differential phase shift is 90º at one frequency in the common passband, it remains close to 90º everywhere in the common passband.

The TL equivalent circuit of the N-layers ML polarizer excited by an incident wave along *y*-axis (Figure 1) can be represented as shown in the Figure 3 where η_0 is the characteristic impedance of the vacuum ($\eta_0 = \sqrt{\mu_0/\varepsilon_0}$) with μ_0 and ε_0 the permeability and the permittivity of vacuum, respectively. The parameters h_{d1}, h_{d2}, …, h_{dN} and $\varepsilon_{r_{d1}}$, $\varepsilon_{r_{d2}}$, …, $\varepsilon_{r_{dN}}$ are respectively the thickness and the relative permittivity of the 1st, 2nd, …, *N*th layer. The parameters h_{s1}, h_{s2}, …, h_{sN} and $\varepsilon_{r_{s1}}$, $\varepsilon_{r_{s2}}$, …, $\varepsilon_{r_{sN}}$ are respectively, the thickness and the relative permittivity of the 1st, 2nd, …, *N*th separator. The L_{eq_1}, L_{eq_2}, …, L_{eq_N} and C_{eq_1}, C_{eq_2}, … , C_{eq_N} are respectively, the lumped inductance and capacitance equivalents of the 1st, 2nd, …, *N*th periodic strip lines on each layer.

The input parameters of the TL model of ML polarized are: the frequency bandwidth (center frequency f_0), the number of layers and the relative permittivity and thickness of each dielectric substrate and separator. This model must be solved by optimizing the shunt components L_{eq_1}, L_{eq_2}, …, L_{eq_N} and C_{eq_1}, C_{eq_2}, …, C_{eq_N} which verify the maximum transmission coefficients (S_{21} and S_{43}) in term of magnitude (~0dB) while

attaining 90^0 differential phase shift between both components for the CP properties.

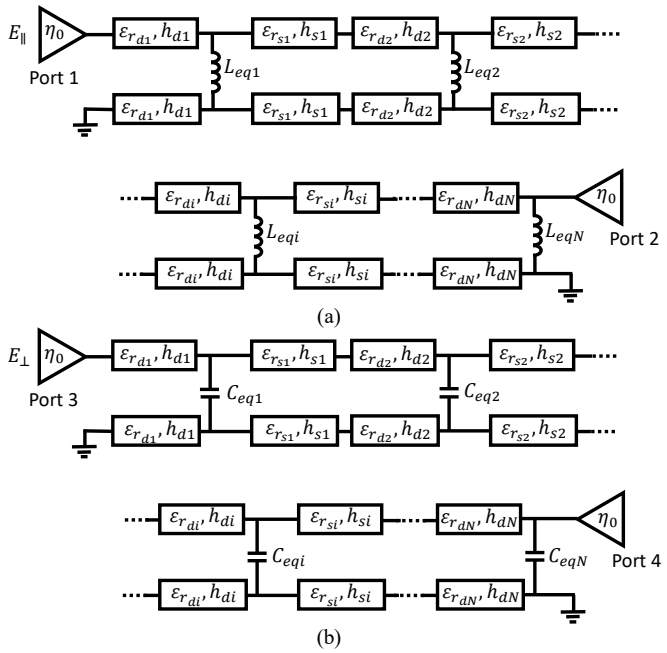


Figure 3: Transmission Line equivalent circuit for N-Layers ML polarizer observed by (a) the parallel wave $\boldsymbol{E}_{\parallel}$ and (b) the perpendicular wave $\boldsymbol{E}_{\perp}$.

3.2. Simulated Annealing Algorithm

Simulated Annealing is a mathematical equivalent of the thermodynamic annealing. The energy of the particle in thermodynamic annealing process can be compared with the cost function to be minimized in optimization problem. The particles of the solid can be compared with the independent variables used in the minimization function. The Simulated Annealing Algorithm is detailed in [15].

The MLs dimensions must be optimized to verify that the impedance equivalent of each layer is equal to the one found by solving the TL circuit model.

Chu and Lee in [14], presented rigorous empirical formulas of equivalent admittances for parallel ($Y_{\parallel}$) polarization $\boldsymbol{E}_{\parallel}$ and perpendicular ($Y_{\perp}$) polarization $\boldsymbol{E}_{\perp}$ of the ML, dependent on the dimensions a, b, h, and w (Figure 4).

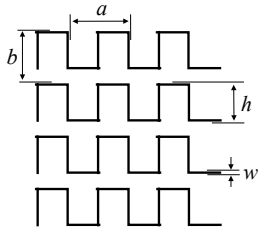


Figure 4: Polarizer's dimensions

The SA algorithm is used to optimize the best values of the ML dimensions of each layer i by making the admittances values ($1/j\omega L_{eq_i}$ and $j\omega C_{eq_i}$), found using the TL circuit model, equal to the ones given by Chu and Lee ($Y_{\parallel_i}(a_i, b_i, h_i, w_i)$ and $Y_{\perp_i}(a_i, b_i, h_i, w_i)$). In other word, for each layer i, the dimensions a_i, b_i, h_i, w_i are optimized to minimize the difference between the admittances ($\delta_i \approx 0$ and $\delta_i' \approx 0$) as shown in the equation (1).

$$\begin{aligned} &\left|Y_{\parallel_i}(a_i, b_i, h_i, w_i) - \frac{1}{j\omega L_{eq_i}}\right| \le \delta_i \\ &\left|Y_{\perp_i}(a_i, b_i, h_i, w_i) - j\omega C_{eq_i}\right| \le \delta_i' \end{aligned} \quad \text{for } i = \{1, \dots, N\}. \quad (1)$$

The optimization procedure is summarized in Figure 5.

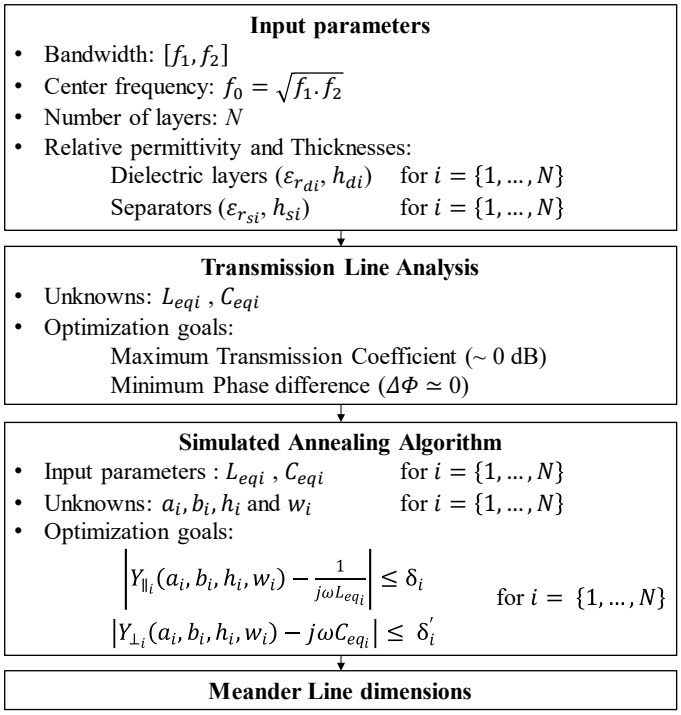


Figure 5: Optimization Procedure Summary

4. Case Study: 6-18GHz Four Layers ML Polarizer

A four layers ML polarizer covering the bandwidth from 6 GHz to 18 GHz is studied in this section.

4.1. Polarizer Optimization

The optimization procedure starts by defining the input parameters which are the frequency bandwidth [6-18GHz], center frequency f_0 = 10.4GHz, number of layers N = 4, relative permittivity and thickness of the dielectric substrates ($\varepsilon_{r_{di}}, h_{di}$) and separators ($\varepsilon_{r_{si}}, h_{si}$) for $i = \{1, \dots, 4\}$.

To simplify the TL equivalent model, the equivalent circuit was studied with one-quarter wavelength apart ($\lambda/4$) air separators and without considering the substrate effect since a very thin substrate ($h_{di} = 0.127mm$) with low permittivity and low loss (Rogers RT/duroid 5880 ($\varepsilon_{r_d} = 2.2, Tan\delta_d = 0.0009$)) is used.

The choice to have the same MLs dimensions (i.e. same L_{eq} and C_{eq}) for each layer was also made to make the system integration simple and flexible. The TL equivalent circuit of the four layers ML polarizer is shown in the Figure 6.

The optimization tool of ADS is used to find the best values of the equivalent inductance and capacitance to verify the maximum transmission coefficients (S_{21} and S_{43}) in term of magnitude (~ 0dB) and 90^0 differential phase shift between them

over the entire bandwidth. The optimum values are: C_{eq}=3 fF and L_{eq} = 0.494 pH.

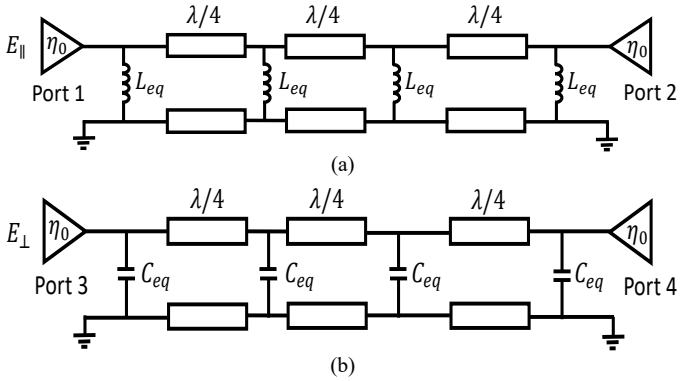
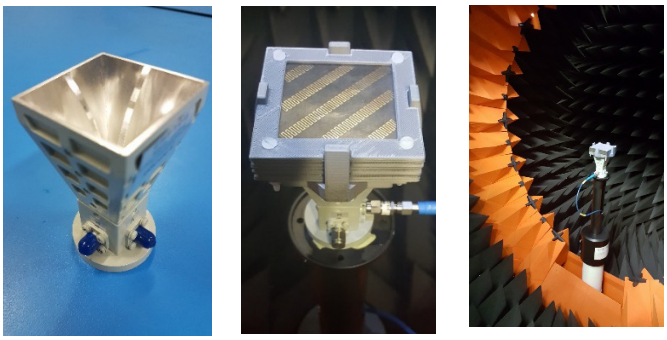


Figure 6: TL circuit model of the four layers Meander Line polarizer observed by the (a) parallel wave ($\boldsymbol{E}_\parallel$) and the (b) perpendicular wave ($\boldsymbol{E}_\perp$).

The ML dimensions (a, b, h, and w) are optimized using the SA algorithm, as detailed previously, by making the admittances values ($1/j\omega L_{eq}$ and $j\omega C_{eq}$), found using the TL circuit model (Figure 6), as close as possible to the ones given by Chu and Lee (1). The optimum values found are a = 2.17mm, b = 13.08mm, w = 0.446mm and h = 5.151mm.

4.2. Polarizer test with 6-18GHz Horn Antenna

To verify the performance of ML polarizer, a wideband horn antenna (6 to 18 GHz) shown in figure 7(a) is used. The four ML layers of the polarizer were fabricated. A mechanical support was designed and realized using a 3D printer to maintain the quarter wave length distance between each layers and the antenna. The picture of the polarizer attached to the antenna is shown in figure 7(b). Figure 7(c) shows the antenna with the ML polarizer under test in the near field anechoic chamber StarLab (by Microwave Vison Group), available in IETR.



(a) (b) (c)

Figure 7: Picture of (a) horn alone - (b) horn with ML-Polarizer – (c) horn with ML-Polarizer under test

The measurements were performed for the horn antenna alone and with the ML polarizer excited in vertical polarization. The reflection coefficients of the horn antenna alone (without the ML polarizer) and with the ML polarizer are plotted in Figure 8.

Both reflection coefficients are very close, which proves that the ML polarizer doesn't mismatch the antenna except at the around 6GHz, where the reflection coefficient is measured around -6.5dB.

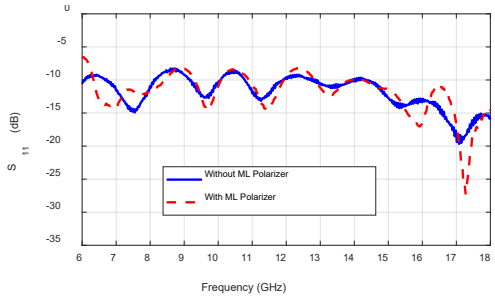


Figure 8: Reflection coefficients of the horn antenna with and without the ML polarizer

The co-polarization and cross-polarization of the Electrical filed (E_θ and E_ϕ) are plotted for different frequencies in the figures 9-15.

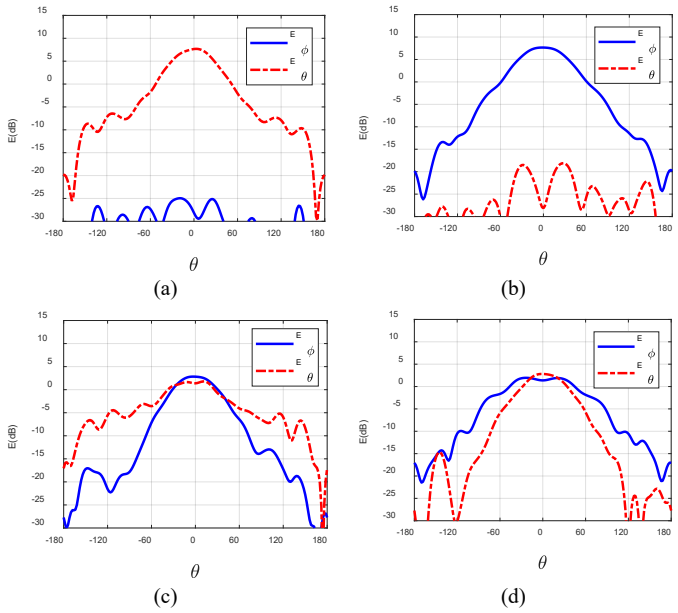


Figure 9: Measured E_θ and E_ϕ at 6 GHz for the horn antenna alone in both planes (a) $\phi = 0^0$ and (b) $\phi = 90^0$ and for the horn antenna with the ML polarizer in both planes (c) $\phi = 0^0$ and (d) $\phi = 90^0$.

In figure 9, the E_θ and E_ϕ components of horn antenna alone and with ML polarizer are plotted at 6 GHz in both planes ($\phi = 0^0$ and $\phi = 90^0$). The results seem to show that the horn antenna initially excited for a vertical polarization (Figure 9 (a) and (b)) is radiating a circular polarized wave when it is stacked with the ML polarizer (Figure 9 (c) and (d)). At the boresight, the difference between the co and cross components of the horn antenna measured alone is more than 28 dB while it reduces to 1dB when the horn antenna is measured with the ML polarizer. The measured Axial Ratio (AR) of the horn antenna with the ML polarizer at this frequency is around 6dB (Figure 16). Therefore, the ML polarizer is converting the linear polarization waves into elliptical polarized ones. The measured insertion losses are around 2.5dB (Figure 17) which is due to the mismatch of the antenna with the ML polarizer at this frequency (Figure 8).

The E_θ and E_ϕ components of the horn antenna alone and with ML polarizer are plotted in figure 10 – 14 for the frequencies from 8GHz to 16GHz. The vertical polarized waves of the horn antenna alone are perfectly converted to circularly polarized ones when it is stacked to the ML polarizer with the maximum AR of

2.9 dB at 8GHz (Figure 16). The insertion losses are less than 1.4 dB with the minimum of 0.24 dB at 14 GHz (Figure 17).

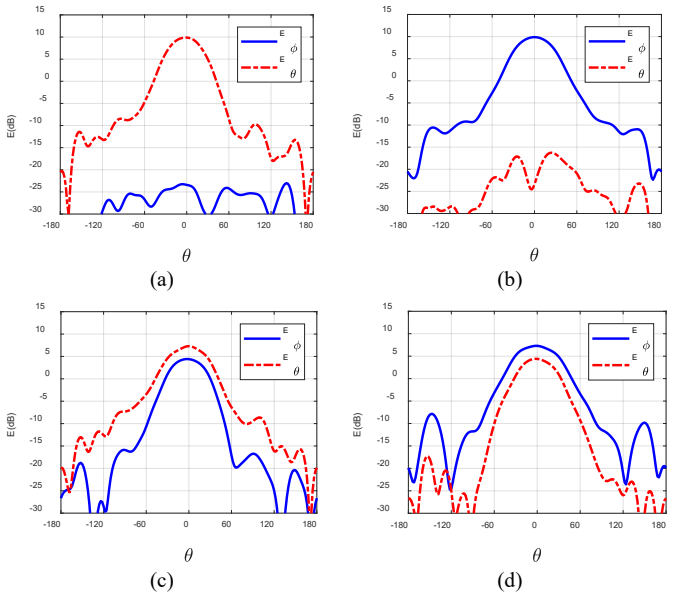


Figure 10: Measured E_θ and E_ϕ at 8 GHz for the horn antenna alone in both planes (a) $\phi = 0^0$ and (b) $\phi = 90^0$ and for the horn antenna with the ML polarizer in both planes (c) $\phi = 0^0$ and (d) $\phi = 90^0$.

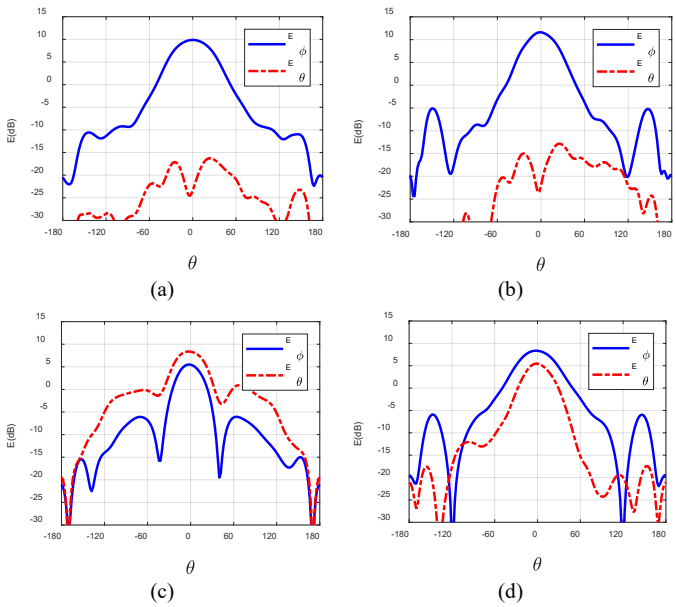


Figure 11: Measured E_θ and E_ϕ at 10 GHz for the horn antenna alone in both planes (a) $\phi = 0^0$ and (b) $\phi = 90^0$ and for the horn antenna with the ML polarizer in both planes (c) $\phi = 0^0$ and (d) $\phi = 90^0$.

The measured results at 18 GHz are plotted in figure 15. At the plane $\phi = 0^0$, some ripples are observed in the pattern of the co-component when the antenna is measured with the ML polarizer. This comportment is not reproduced at the other cut plane ($\phi = 90^0$) where the directive radiation pattern shape is preserved. The measured Axial Ratio (AR) of the horn antenna with the ML polarizer at this frequency is around 5.4 dB (Figure 16). Therefore, the linearly polarized waves of the horn antenna alone are converted to elliptical polarized waves. The measured insertion losses are around 0.6 dB (Figure 17).

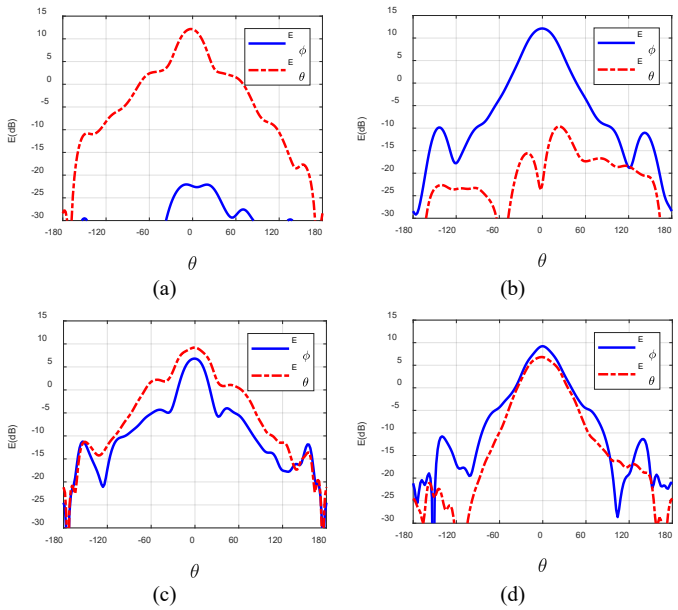


Figure 12: Measured E_θ and E_ϕ at 12 GHz for the horn antenna alone in both planes (a) $\phi = 0^0$ and (b) $\phi = 90^0$ and for the horn antenna with the ML polarizer in both planes (c) $\phi = 0^0$ and (d) $\phi = 90^0$.

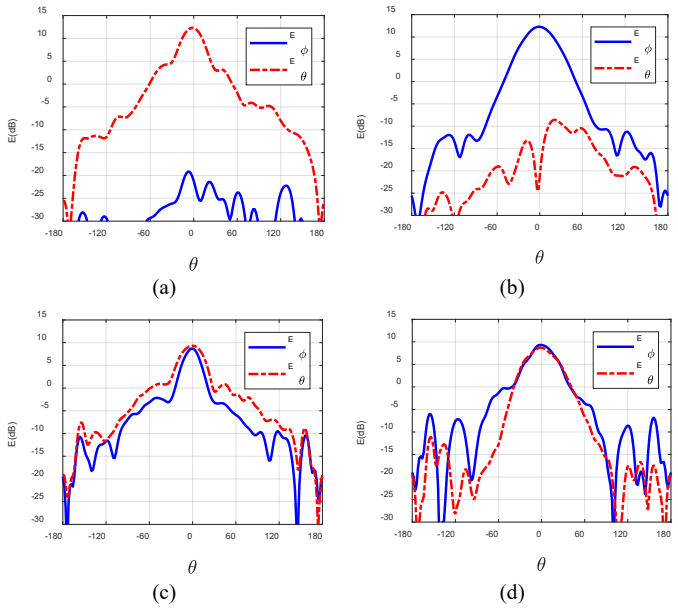


Figure 13: Measured E_θ and E_ϕ at 14 GHz for the horn antenna alone in both planes (a) $\phi = 0^0$ and (b) $\phi = 90^0$ and for the horn antenna with the ML polarizer in both planes (c) $\phi = 0^0$ and (d) $\phi = 90^0$.

The measured values of the AR at boresight are reported in figure 16 versus frequency for the horn antenna with the ML polarizer. The AR is less than 3dB at the frequency band between 8 GHz and 16 GHz. Therefore, the ML polarizer converts the linear polarization of the horn antenna to a circular one. In both ends of the frequency band (6 GHz and 18 GHz), the ML polarizer converts the linear polarization of the horn antenna to an elliptical one.

The total electrical fields (E_{tot}) values at boresight are reported in figure 17 versus frequency for the horn antenna alone and with the ML polarizer. This figure shows clearly that the E_{tot} of the horn

antenna with the ML polarizer is following the same allure as the E_{tot} of the horn antenna alone variation with some insertion losses.

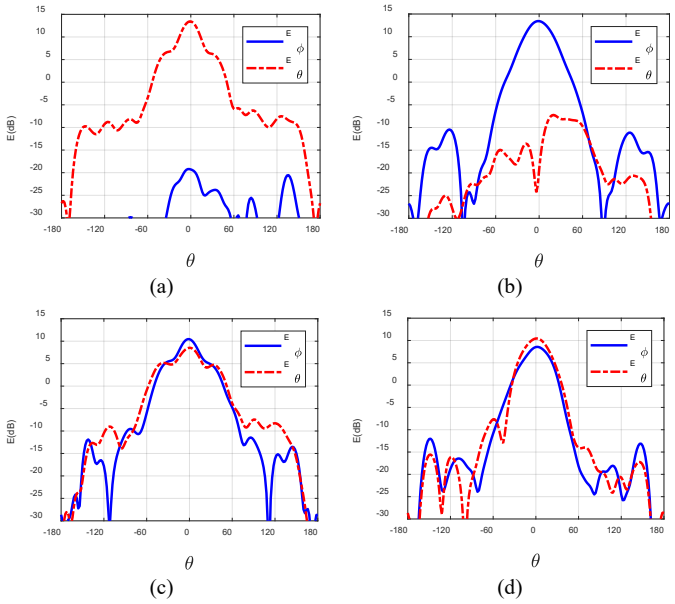


Figure 14: Measured E_θ and E_ϕ at 16 GHz for the horn antenna alone in both planes (a) $\phi = 0^0$ and (b) $\phi = 90^0$ and for the horn antenna with the ML polarizer in both planes (c) $\phi = 0^0$ and (d) $\phi = 90^0$.

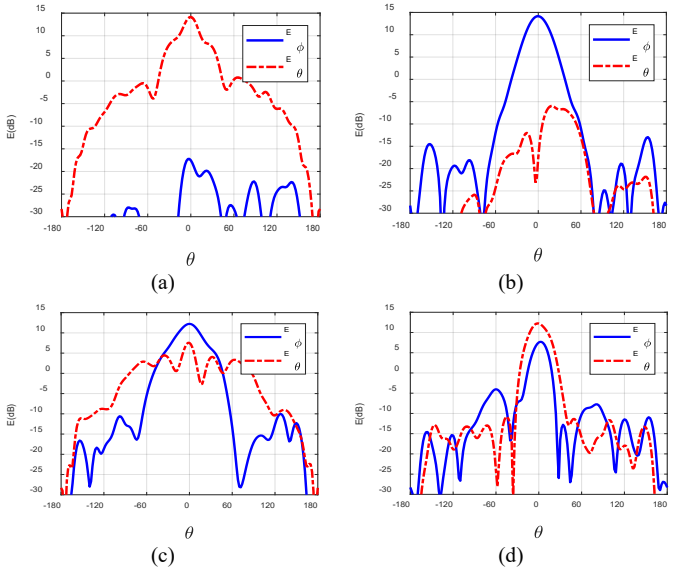


Figure 15: Measured E_θ and E_ϕ at 18 GHz for the horn antenna alone in both planes (a) $\phi = 0^0$ and (b) $\phi = 90^0$ and for the horn antenna with the ML polarizer in both planes (c) $\phi = 0^0$ and (d) $\phi = 90^0$.

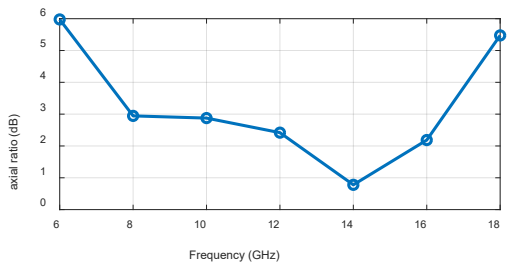


Figure 16: Measured Axial Ratio (***AR***) versus frequency for the horn antenna with the ML polarizer

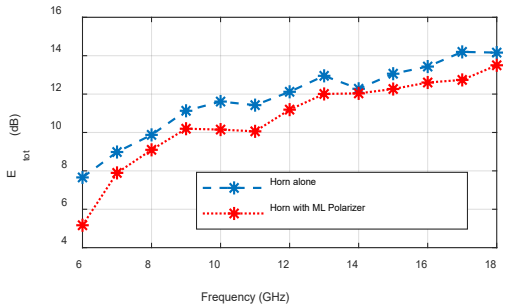


Figure 17: Measured total electrical field (***E_{tot}***) versus frequency for the horn antenna alone and with the ML polarizer

In almost the entire bandwidth, except at 6 GHz, the insertion losses are less than 1.5dB with the minimum of 0.24 dB at 14 GHz. The maximum insertion losses (2.5dB) are measured at 6 GHz which is due to the mismatch of the horn antenna with the ML polarizer.

5. Conclusion

An optimization Method of Wideband Multilayer Meander-Line Polarizer using a Semi-Analytical approach is described in this paper. The originality of this method lies in the optimization of: i) values of the shunt equivalent inductance L_{eq} and capacitance C_{eq}, determined by the ADS software - ii) ML dimensions (a, b, h and W) using the SA algorithm. This method was applied to optimize a four layers wideband (6 – 18 GHz) ML polarizer as study case.

The optimized polarizer was manufactured and a wideband Horn antenna covering the same bandwidth was used for the testing. The test consists of the measurements of the horn antenna vertically polarized without and with the ML polarizer. Good measured results were observed where the linear polarization of the antenna was converted to a purely circular one in almost the entire frequency band (8 GHz – 16 GHz, corresponding to an octave). In both ends of the frequency band (6 GHz and 18 GHz), an elliptical polarization was achieved with a maximum axial ration of around 6dB.

At 6GHz, the axial ratio is around 1dB but the insertion losses are around 2.5dB which is due to the mismatch of the horn antenna with the ML polarizer at this frequency. At 18 GHz, the axial ratio is around 5dB and the insertion losses are around 0.6 dB.

The future work will be devoted to the study of five layers ML polarizer to improve the performances, and especially at both ends of the frequency band.

References

[1] W. Abdouni, M. S. Khan, A. Konstantinidis "Design of a Wideband Multilayer Meander-Line Polarizer (6 – 18 GHz) using a Semi-Analytical Method," in 48th European Microwave Conference (EuMC), Madrid Spain, 2018, https://doi.org/10.23919/EuMC.2018.8541675

[2] R. A. Marino, "Accurate and efficient modeling of meanderline polarizers," Microw. J., 41 (11), 22-34, Nov. 1998

[3] J. F. Zurcher, "A meander-line polarizer covering the full E-band (60-90 GHz)," Microwave Opt. Technol. Lett., 18, 320-323, Aug. 1998. https://doi.org/10.1002/%28SICI%291098-2760%2819980805%2918%3A5<320%3A%3AAID-MOP3>3.0.CO%3B2-A

[4] L. Young, L.A. Robinson, and C.A. Hacking "Meander-Line Polarizer," IEEE Transactions on Antennas and Propagation, 21(3), 376-378, 1973. https://doi.org/10.1109/TAP.1973.1140503
[5] J. J. Epis, "Broadband antenna polarizers," US Patent 3754271, Aug. 1973.
[6] T. L. Blackney, J. R. Burnett, and S. B. Cohn, "A design method for meander-line circular polarizers," 22nd Ann. Antenna Symp., 1-5, Oct. 1972.
[7] C. Terret, J. R. Level, and K. Mahdjoubi, "Susceptance computation of a meander-line polarizer layer," IEEE Transactions on Antennas and Propagation, 21, 1007-1011, Sep. 1984. https://doi.org/10.1109/TAP.1984.1143456
[8] S. W. Lee, and G. Zarrilo, "Simple formulas for transmission through periodic metal grids or plates," Antenna Application Sym., University of Illinois, Urbana, IL, Sponsored by RADC (EEA) Electromagnetic Sciences Division, Hanscom AFB, 1983. https://doi.org/10.1109/TAP.1982.1142923
[9] A. K. Bhattacharyya, and T. J. Chwalek, "Analysis of multilayered meander line polarizer," Int. J. of Microwave and Millimeter-wave Computer-aided Eng., 442-454, 1997. https://doi.org/10.1002/(SICI)1522-6301(199711)7:6%3C442::AID-MMCE6%3E3.0.CO;2-P
[10] M. A. Joyal, and J.-J. Laurin, "Analysis and design of thin circular polarizers based on meander lines," IEEE Transactions on Antennas and Propagation, 60, 3007-3011, Sep. 2012. https://doi.org/10.1109/TAP.2012.2194659
[11] C. C. Chen, "scattering by a two-dimensional periodic array of conducting plates," IEEE Transactions on Antennas and Propagation, 18, 660-665, 1970. https://doi.org/10.1109/TAP.1970.1139760
[12] R. S. Chu, and K. M. Lee, "Radiation impedance of a dipole printed on periodic dielectric slabs protruding over a ground plane in an infinite phased array," IEEE Transactions on Antennas and Propagation, 35, 13-25, 1987. https://doi.org/10.1109/TAP.1987.1143966
[13] M. Letizia, B. Fuchs, C. Zorraquino, J.-F. Zurcher, and J. R. Mosig, "Oblique incidence design of meander-line polarizers for dielectric lens antennas," Progress In Electromagnetics Research B, 45, 309- 335, 2012. https://doi.org/10.2528/PIERB12080807
[14] RUEY-SHI CHU, KUAN-"I LEE, "Analytical Model of a Multilayered Meander-Line Polarizer Plate with Normal and Oblique Plane-Wave Incidence," IEEE Transactions on Antennas and Propagation, 35(6), 652 – 661, 1987. https://doi.org/1987 10.1109/TAP.1987.1144158
[15] E.S. Gopi, "Algorithm Collections for Digital Signal Processing Applications Using Matlab", Mathematics\\Wavelets and signal processing, 2007.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 139-142 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Development of Teaching Aids Selection Standards in Maker Education

Woochun Jun*

Seoul National University of Education, Department of Computer Education, 06639, Korea

ARTICLE INFO

Article history:
Received: 30 May, 2019
Accepted: 28 August, 2019
Online: 24 September, 2019

Keywords:
Maker Education
Selection Standards
Teaching Aids
Physical Computing

ABSTRACT

Due to the recent development of information and communication technology, such as sensor technology, IoT technology, and smart technology, interest in maker education is increasing. Maker education offers students experience-based education that allows them to make their own results, giving them various interests and motivations. Recently, various research results on maker education have been under way. However, there is little research work on the teaching aids selection criteria for maker education. The purpose of this study is to justify the teaching aids selection standards that are developed in the previous research works. In the previous research work, 14 selection standards were developed in a total of eight areas. In this work, through rigorous statistical analysis, justification of the 14 standards are verified. The findings are expected to help a lot in future policy proposals and related research on maker education.

1. Introduction

This paper is an extension of work originally presented in ICTC2018[1].

Various changes are taking place throughout the industry following the recent advent of the 4th industrial revolution era. Information-oriented industries are emerging at the forefront of the industry in the existing product-oriented manufacturing industry structure. The changes are prompting the acceleration of the modern knowledge and information society, calling for, among other things, interest in and understanding of the information-oriented society.

This change in the industrial structure also calls for many changes in public education in schools. In particular, interest in information education such as information collection, analysis, and processing is increasing in schools as the knowledge and information society accelerates. Information education is evolving into a software education form, especially from Information and Communication Technology Education, which started from 2000 in Korea. While the human image sought in ICT education was the core part of a computer-savvy user, the human image sought by the government in 2015 became the producer or manufacturer who could actually make software [2].

Maker education began in the early 2000s with the U.S. as the Do-It-Yourself Movement [3]. The DIY movement aimed to produce and use the necessary items in everyday life and also to share the necessary processes and products with each other. The DIY movement evolved into the Maker Movement, which, combined with the Maker Space, was ultimately born into Maker Education [4].

Meanwhile, according to [5], maker education was defined as follows by relating it to information education (or software education). In [5], maker education is a new type of educational paradigm to foster so-called "maker geeks" who can create things themselves. In other words, what students have imagined autonomously is produced directly using digital devices. It was also defined as process-oriented project education rather than the results that lead to the sharing of knowledge and experience learned in this process.

Maker education is in the spotlight as an experience-type education that can increase students' interest and motivation for learning. It is also receiving attention at the schools in a typical form of convergence education, combined with various subjects, such as science and practical subjects. In addition, various forms of hands-on exhibitions and contests are attracting interest from students and parents, and teachers are given various training opportunities to encourage interest in maker education.

Along with the spread of maker education, academic research is under way on various maker education. However, in order to carry out actual maker education on the schools, there is a lack of standards on how to choose teaching aids, which is the core of the maker education, among other things Generally, teaching aids are

*Corresponding Author: Woochun Jun, Seocho Dong, Seocho Gu, Seoul, Korea, wocjun@snue.ac.kr

tools used in class activity. The typical teaching aids in class are scissors, colored paper, glue, and so on. In maker education, the typical teaching aids includes 3D printers, 3D pens, cutting tool, sewing machine, and so on. Teaching aids are either raw materials or making tools like a machine tool. In maker education, teaching aids are essential for class activity. Thus, choosing right teaching aids in maker education is very important in successful class.

The purpose of this study is to justify teaching aids selection standards that are developed in the previous research works. That is, for those 14 selection standards, rigorous statistical work is done and selection standards are proved to be useful in schools.

The rest of this paper is as follows. First, in Chapter 2, related works are presented. In Chapter 3, the existing 14 selection standards are introduced. In Chapter 4, statistical analysis is presented for selection standards. Finally, in Chapter 5, conclusions and further research issues will be discussed.

2. Related Works

This chapter presents an overall relevant study of maker education. First, the typical characteristics of maker education can be summarized as follows [6].

First, knowledge can be organized through interaction with objects or people through direct-creating activities. In addition, outputs created directly by learners can induce social participation through various sharing, thereby providing a meeting place for various participants to communicate together through common tasks.

Second, it can provide self-directed and cooperative learning. Makers are engaged in maker activities, choosing their own themes from areas of personal interest and social issues. In addition, working on cooperative projects can lead to mutual respect and communication through cooperative learning.

Third, it provides an environment for making activities. In other words, the maker education will require a maker space for direct production activities and a variety of teaching aids and materials. The maker space needs an environment with the tools needed for the maker's activities.

Fourth, maker education requires a position as an inventor and creator as a learner. In other words, students go through the process of choosing their own themes through making activities. Through this process, learners experience creative problem solving.

Fifth, in maker education, a teacher will act as a helper. Maker activities are basically activities that can improve students' creative thinking and problem-solving skills. In addition, since students decide on their own topics, tools, methods, etc., they should provide a free learning environment and a free environment for evaluation regardless of the outcome's success or failure. Therefore, a teacher's role should be a learning guide or facilitator.

Sixth, students can start with personal participation in maker education and continue with social participation. This is because the maker education can share and open information with other learners through personal problem solving, starting from social issues and creating experiences, and to see if personal and social problems can be solved through product sharing.

On the other hand, a recent study by [7] developed and presented a measure of the ability of making as part of the software coding- based maker education program for scientific gifted children. The Matching Capacity Scale means the competency that students who participate in the maker education should have, and the makeup competency measures presented in [7] are as follows. Overall, the making capabilities comprise three core competences: analysis capabilities, design capabilities, and implementation capabilities. Specifically, analysis capability refers to the ability to collect data and analyze problems, and design capability corresponds to the ability to generate and refine ideas for problem resolution, and to come up with solutions through sharing ideas. Finally, implementation capability refers to the ability to select a tool for problem solving, generate results through active participation and evaluate results. The three core competences were divided into more detailed sections to provide specific capabilities.

Table 1. The Original Selection Standards

Number	Standard	Detailed Criteria
1	Safety	-There should be no risk of safety accidents. -Material that can be harmful to health should not be used
2	Subject Compatibility	-It should be available for various educational activities -It should be possible to integrate with other teaching materials and learning materials.
3	Multi-purpose	-The teaching aids must support various functions through the connection of built-in or additional parts.
4	Ease of Manipulation	-(Teacher) It should be easy to use in class by simple training. -(Student) It should be easy to use after simple teaching and learning about how to use.
5	Economics	-It should be easy to buy within budget items of the school. -Parts can be easily procured within reasonable time and price.
6	Universal Design	-It should be designed for the older, people with mental and physical disabilities. -It should provide alternative functions for main functions
7	Gender Equality	-It should be used regardless of gender.
8	Durability	-Failure or breakage should not occur even if students repeatedly use. -It should be able to withstand the minor impact (collision, drop, etc.) that occurs between educational activities.

3. Development of Teaching Aids Selection Standards

3.1 Design Principles

The selection standards for teaching aids in maker education are developed through the following principles [1].

First, it is in line with the general teaching aids selection criteria used in primary and secondary schools. That is because the teaching aids used in maker education are essentially one of the teaching aids.

Second, it is based on the selection criteria for the physical computing education, which is mainly used in recent software education.

Third, the universal design principle is adopted so that it can be used by anyone. In other words, care is taken to make it easier for anyone, regardless of age, gender, age, ordisability.

3.2 The Original Selection Standards

In [1], the original selection standards of teaching aids in maker education are proposed. The following table 1 show the initial standards.

4 Statistical Analysis

4.1 Design of Analysis

Verification of the reliability of the tool to measure the importance of teaching aids selection standards in maker education showed that Cronbach α was 0.77 and was reliable, as shown in Table 2.

Table 2. Reliability of Measurement Tool

Item Number	Cronbach's α
14	0.77

The data gathered in this work were analyzed using the Statistical Package for the Social Science (SPSS) WIN 25.0 program. Cronbach α is used to verify the reliability of the measuring tool. In addition, the average and standard deviation were calculated in order to find out the importance of the selection standards in maker education, and the correlation analysis was conducted to understand the relationship between the sub-areas of selection standards and selection standard in maker education.

4.2 Analysis Results

-Descriptive Statistics

The average and standard deviation of questions asked about the importance of teaching aids selection standards in maker education are as shown in Table 3.

4.3 Sub-area of Descriptive Statistics

According to the results of the review of the average and standard deviation in the sub-areas of teaching aids selection standards in maker education are shown in table 4. The average importance of the selection standards for the teaching aids in maker education was 4.13, and the selection standards for the teaching aids were important.

Among the sub-areas of the standards for selection standards for teaching aids, “safety” 4.84, followed by "ease of manipulation" 4.52, "durability" 4.30, "economics" 4.28,"subject compatibility" 4.05, "gender equality" 4.00 and "multi-purpose" 3.78, with the lowest "universal design."

As above, it is shown that “safety” is the highest in the sub-areas of the teaching aids selection standards in maker education, and that “universal design” is less important than other areas.

Table 3. Results of Descriptive Statistics

Standard	AVG	SD
1. There should be no risk of safety accidents.	4.84	0.37
2. Material that can be harmful to health should not be used.	4.84	0.37
3. It should be available for various educational activities	4.13	0.91
4. It should be possible to integrate with other teaching materials and learning materials.	3.97	0.82
5. The teaching aids must support various functions through the connection of built-in or additional parts.	3.78	0.94
6. (Teacher) It should be easy to use in class by simple training.	4.41	0.80
7. (Student) It should be easy to use after simple teaching and learning about how to use.	4.63	0.61
8. It should be easy to buy within budget items of the school.	4.38	0.66
9. Parts can be easily procured within reasonable time and price.	4.19	0.86
10. It should be designed for the older, people with mental and physical disabilities.	3.06	1.16
11. It should provide alternative functions for main functions	3.00	1.14
12. It should be used regardless of gender.	4.00	1.19
13. Failure or breakage should not occur even if students repeatedly use.	4.28	0.85
14. It should be able to withstand the minor impact (collision, drop, etc.) that occurs between educational activities.	4.31	0.69

Where AVG and SD stand for average and standard deviation, respectively.

4.4 Correlation Analysis

The results of a review of the correlation between the sub-areas of the standards for teaching aids selection standards in maker education and the whole teaching aids selection standards in maker education are shown in Table 5.

Safety (r=.455; p<.01) and subject compatibility (r=.627, p<.001), multi-purpose (r=.548, p<.01), ease of manipulation (r=.391, p<.05), economics (r=.534, p<.01), universal design (r=.725, p<.001), gender equality (r=.824, p<.001), and durability (r=.453, p<.01) showed a statistically significant static correlation with the importance of the standards for teaching aids selection

standards in maker education. Therefore, the higher importance of safety and subject compatibility, multi-purpose, ease of manipulation, economics, universal design, gender equality, and durability of the standards for teaching aids selection standards in maker education, the higher importance of the standards for teaching aids selection standards in maker education.

Table 4. Results of Descriptive Statistics of Sub-area

Evaluation Standard Area	AVG.	SD
Safety	4.84	0.27
Subject Compatibility	4.05	0.73
Multi-purpose	3.78	0.94
Ease of Manipulation	4.52	0.63
Economics	4.28	0.61
Universal Design	3.03	1.07
Gender Equality	4.00	1.19
Durability	4.30	0.69
Overall Selection Standards	4.13	0.43

where AVG and SD stand for average and standard deviation, respectively.

Table 5. Results of Correlation Analysis

Area	Correlation Value
Safety	0.455**(0.009)
Subject Compatibility	0.627***(0.000)
Multi-purpose	0.548**(0.001)
Ease of Manipulation	0.391*(0.027)
Economics	0.534**(0.002)
Universal Design	0.725***(0.000)
Gender Equality	0.824***(0.000)
Durability	0.453**(0.009)

Where * p<.05, ** p<.01, *** p<.001

5 Conclusions and Further Research Issues

Along with the development of various information and communication technologies and smart technologies, maker education is receiving attention at schools due to the increase in the number of available teaching aids at schools. Maker education is an experience-based education in which students directly participate in designing and analyzing the work they want to produce, and finally participate in production, which has the advantage of generating interest and motivation for students. Maker education is also very popular as a project-oriented cooperative study because teachers and students can participate and discuss the entire process of producing works together.

The role of the teaching aids in maker education is very important and the interest and academic performance of students in the whole class, depending on how they choose the teaching aids, also affects safety issues. However, there has been no standardized research or guideline on the standards for teaching aids selection standards in maker education.

The purpose of this study is to justify the teaching aids selection standards that are developed and proposed. In the previous research, 14 selection standards are developed and analyzed statistically. In this research, with more rigorous statistical analysis, the justification of the results is verified.

The immediate further research issues are as follows. At first, it is necessary to present more detailed selection standards depending on tools of maker education such as 3D printer. Also, it is also necessary to present more integrated and extensive selection standards for both maker education and software education. This is because there is no clear limit between maker education and software education.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] W. C. Jun, "A Study on Development of Selection Standards of Maker Education Teaching Aids" in 2018 International Conference on Information and Communication Technology Convergence, Jeju Island, Korea, 2018.

http://10.1109/ICTC.2018.8539704

[2] Korea Ministry of Education, "Software Education Training and Operation Guidelines", 2015.

[3] E. R. Halverson, and E. M. Sheridan, "The Maker Movement in Education", Harvard Business Review, 84(4), 495-504, 2014.

https://doi.org/10.17763/haer.84.4.34j1g68140382063

[4] D. Dougherty, "The Maker Movement, Innovations: Technology, Governance, Globalization", MIT Press, 7(3), .11-14, 2012.

https://doi.org/10.1162/INOV_a_00135

[5] Seoul Metropolitan Office of Education, "'2018 ~ '2022 Mid and Longterm Maker Education Development Master Plan", November 1, 2017.

[6] M. J. Kang, "Development of Checklist for Development and Operation of Maker Education Program", Master Thesis, Kyunghee University, 2018.

[7] J. H. Lee and J. H. Jang, " Development of Maker Education Program based on Software Coding for the Science Gifted", Journal of Gifted/Talented Education, 27(3) 331-348, 2017.

http://doi.org//10.9722/JGTE.2017.27.331

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 143-150 (2019)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Application of Fractal Algorithms to Identify Cardiovascular Diseases in ECG Signals

Gustavo Gallegos Dávalos, Pedro Freddy Huamaní Navarrete*

Ricardo Palma University, Engineering Faculty, Lima 33, Peru

ARTICLE INFO

Article history:
Received: 27 May, 2019
Accepted: 04 July, 2019
Online: 24 September, 2019

Keywords:
ECG signals
Physionet database
Higuchi and Katz algorithm
fractals
Principal Component Analysis

ABSTRACT

The aim of this article was the identification of cardiovascular diseases, after applying Katz and Higuchi fractal algorithms on 4 databases of ECG signals downloaded from the Physionet website: heart failure (HF), hypertension (H), ischemic heart disease (IHD) and normal sinus rhythm (NSR). For this purpose, initially the ECG signals passed through a filtering stage using a Butterworth high pass filter of order 6 and 0.5Hz of cutoff frequency, in order to cancel variations of the baseline. The fractal algorithms were applied independently for each database. For such application, all signals were standardized with a total of 100,000 data, and for the calculation of each fractal dimension (FD) a frame equal to 10,000 with an overlap of 1,000 was used in a first stage; in a second stage, a frame equal to 1,000 with an overlap of 100 was used. Thus, the results showed that the Higuchi algorithm, in general, has a better performance compared with the Katz algorithm. These results refer to the observation of the variance of the FD averages, which are shown in Table 4. For example, in the cardiovascular disease Arterial Hypertension, the Higuchi algorithm presented 0.0093 of variance compared with the algorithm of Katz that only reached 0.0352. Complementarily, we used the Principal Component Analysis (PCA) to show graphically the differences between algorithms, so the components related to the Higuchi algorithm were presented in a less dispersed and grouped form, making possible their differentiation, mainly when the second analysis stage was carried out.

1. Introduction

The circulatory system is one of the such systems of the human being that fulfills a specific function; and, where the most important piece of this system is the heart, because it is responsible for pumping the blood to the different organs and systems of the human body in order to oxygenate them.

Due to the agitated life we lead, poor feeding, combined with the consumption of tobacco and alcohol steadily and in some cases certain hereditary ills, they can produce different types of ilness that damage the heart and affect the natural rhythm of its heartbeat. For this reason, there are several alterations about the forms of cardiac rhythms, which vary in intensity and severity. Thus, the purpose of using ECG signals of patients with cardiovascular diseases of higher incidence arises, such as it was stated by the World Health Organization [1], in order to be identified. For this was necessary to use the Physionet database [2], from which a total of 60 records of electrocardiogram signals were downloaded, including the one corresponding to the normal sinus rhythm. Then, for the identification of cardiovascular diseases in digitized ECG signals, we chose a field of mathematics that is dedicated to the study of signals that have a repetitive characteristic of self-similarity, which is called fractals. Likewise, this theory is also used in the analysis of encephalogram signals EEG, like early detection of Alzheimer's disease [3], to visual evoked potential [4], to analyze the EEG brain oscillations during the development of cognitive tasks [5], and to analyze of the bimodal pattern of neuronal activity [6]. Likewise, the principal components analysis theory was used to reduce the dimensionality of the results obtained from the fractal dimensions of Higuchi and Katz. For this, the work of [7] was taken as reference, who used fractal analysis and chaos theory to detect dynamic changes in a group of 13 ECG signals with healthy and unhealthy segments, mainly when RR intervals and the ST segments were analyzed. Likewise, [8] was taken as reference who show the use of a procedure about analysis of fluctuation without tendency based on the theory of fractals, for patients who had cerebrovascular accidents. Also, the work of [9]

* Pedro Freddy Huamaní Navarrete, Email: phuamani@urp.edu.pe

where they indicate the development of a local fractal characteristic based on a method of concordance for the classification of arrhythmia ECG, by matching it with the representing templates from other ECG signals with different types of arrhythmias; and for this they used the Euclidean distance. Also, [10] used fractal dimension for automated diagnosis of serious arrhythmias. In the same way, the work of [11] investiged three fractal dimension methods, among them, Higuchi and Katz, plus artificial neural networks, as a methodology to predict sudden cardiac death. And, in [12], the use of two methods for the extraction of characteristics in ECG signals is proposed, where the second method is a fusion of fractal characteristics by stacking.

2. Processing and Fractal Dimension

This section briefly describes the four databases used to subsequently apply the filtering step, followed by the calculation of the fractal dimension of the ECG signals. In addition, this database was seleccioned, because represent to a group of cardiovascular diseases that have great incidence in Peru.

2.1. Physionet Database

There were four databases used in this article, which were obtained from the Physionet website [2]. These are:

- The BIDMC Congestive Heart Failure Database, for severe cases of heart failure. Which used a sampling frequency of 250 Hz, for a DII derivation, and with signals of 900,000 samples.
- The MIT-BIH Normal Sinus Rhythm Database, in the case of healthy people. Which used a sampling frequency of 128 Hz, for a DII derivation, and with signals of 460,800 samples.
- St.-Petersburg Institute of Cardiological Technics 12-lead Arrhythmia Database, for the case of hypertension disease. Which used a sampling frequency of 257 Hz and signals of 462,600 samples.
- The PTB Diagnostic ECG Database for the case of ischemic heart disease. Which used a sampling frequency of 1,000 Hz and signals of 115,200 samples.

In the same way, the PhysioBank ATM tool was used to download the files in *. MAT format, for a total of 15 signals per database in order to have a uniformity in all four cases. Figure 1 shows the temporal representation of an ECG signal in its original form, corresponding to a patient with heart failure.

2.2. Filtering Stage

Filtering stage consists of attenuating the baseline or DC component that alter the ECG signal. According to the American Heart Association in its article Recommendations for the Standardization and Interpretation of Electrocardiograms, recommends the use of filters with cutoff frequencies up to 0.67 Hz or less for linear digital filters with phase distortion equal to zero [13]. Therefore, it was used a Butterworth high pass filter with order 6 and cutoff frequency equal to 0.5 Hz, in order to attenuate the DC component and having a stable phase response. Figure 2 shows the comparison between an original signal from a patient with ischemic heart disease, and its corresponding result after applying the filtering stage.

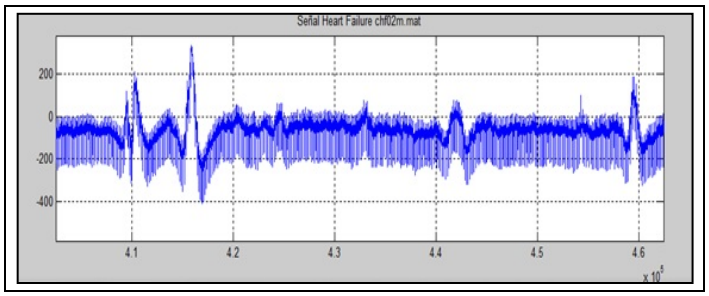


Figure 1. ECG signal from the database BIDMC Congestive Heart Failure to a patient with heart failure. Source: Authors.

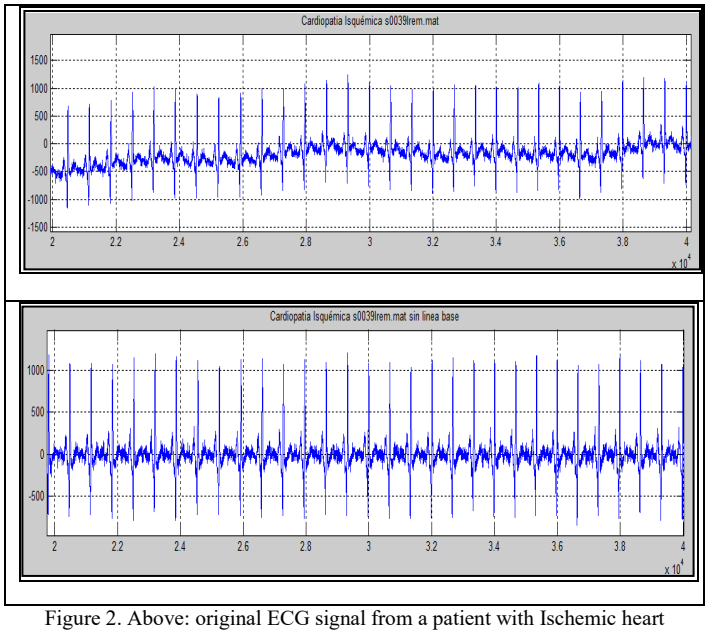


Figure 2. Above: original ECG signal from a patient with Ischemic heart disease. Bottom: ECG signal filtered with Butterworth filter. Source: Authors.

3. Estimation of Fractal Dimensions

The calculation of the fractal dimension of Higuchi considers a finite set of observations of time series taken in a regular interval, to then built a new series of time [14]. It is described using Equation (1).

$$x_m^k = x(m), x(m+k), x(m+2k), \ldots, x\left(m + \left|\frac{N-m}{k}\right| . k\right) \quad (1)$$

For $m = 1, 2, \ldots, k$

Where « m » indicates the initial time value, « k » is the discrete time delay, and $|\alpha|$ means integer part of α.

In this way, algoritm proposed by Higuchi, to estime the fractal dimension, can be seen in Figure 3 as a flow chart**.** This flow chart is based on the program coded by [15].

Likewise, algorithm proposed by Katz to estimate the fractal dimension of a flat curve is shown in Equation (2).

$$\text{Katz's Fractal Dimension(D)} = \frac{\log L}{\log d} \quad (2)$$

Where « d » is the diameter of the curve and « L » is the length of the curve, defined as the distance between two successive points.

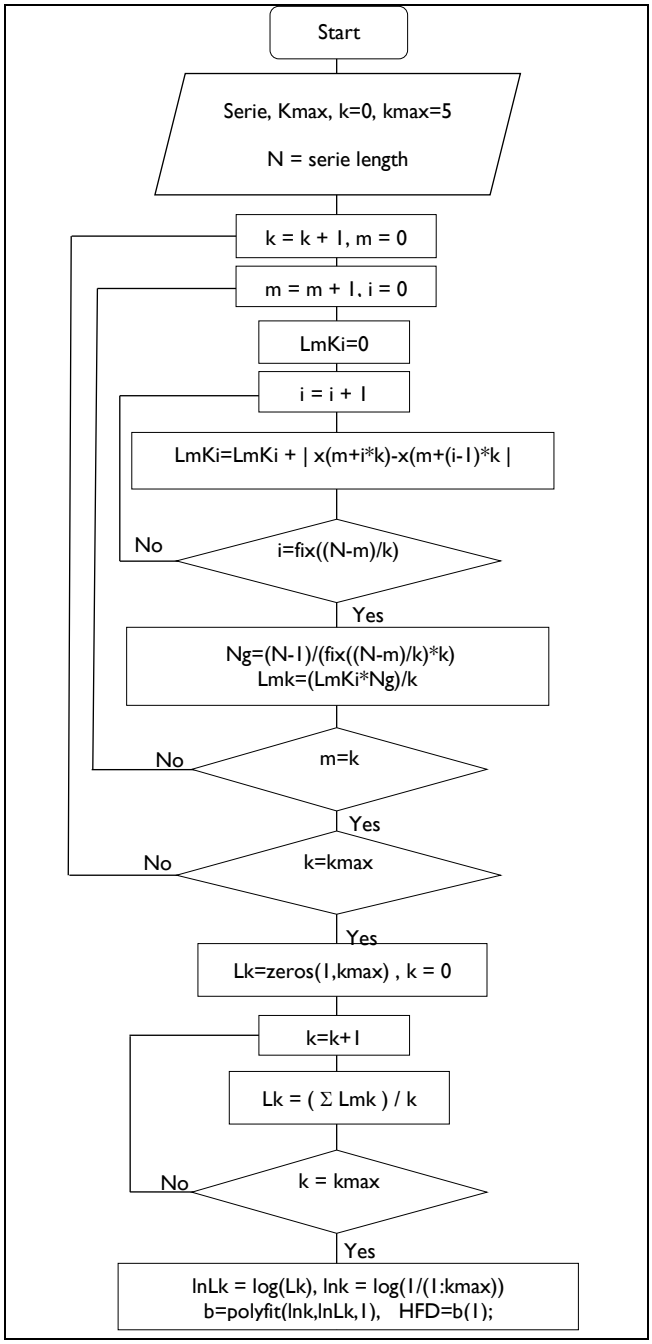


Figure 3. Flow chart to calculate the fractal dimension by Higuchi.
Source: Authors

In addition, Katz adds that by practical convention, in order to discretize the space and to normalize the fractal dimensions, he defines the standard unit as the smallest convolution size of interest in a form, then for waveforms, the average distance between 2 successive points is « a », then using « a » and clearing in the formula for fractal dimension D [16], result Equation (3).

$$\textbf{Katz's Fractal Dimension}(\mathbf{D}) = \frac{\log(n)}{\log\left(\frac{d}{L}\right) + \log(n)} \quad (3)$$

Where « n » is the number of steps in the curve (L / a), and « a » is the average length of the unit. Then, flow chart used to calculate the fractal dimension by Katz can be seen in Figure 4. This flow chart is based on the program coded by [17].

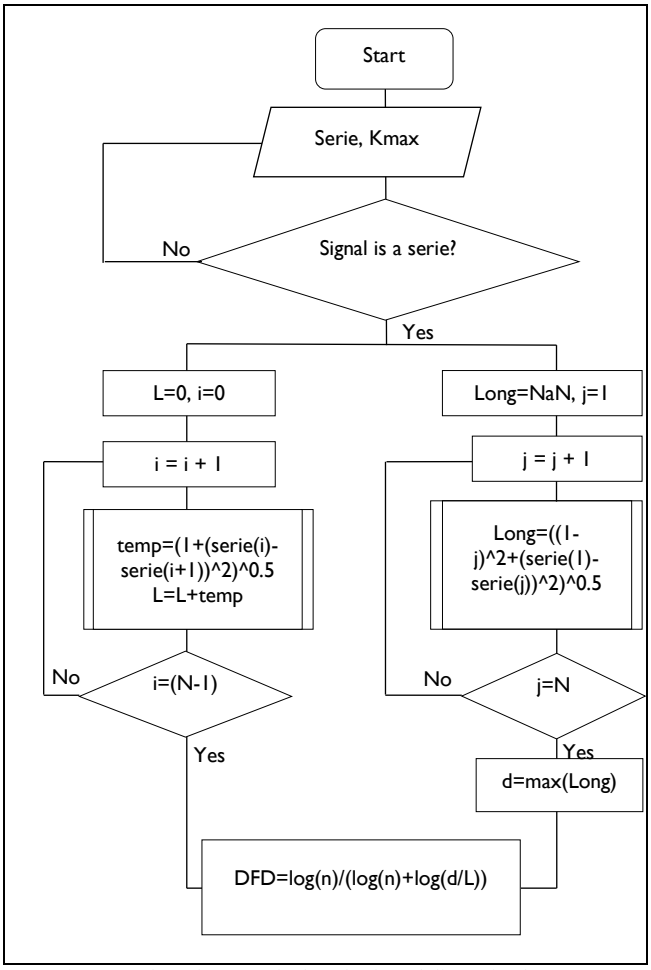


Figure 4. Flow chart to calculate the fractal dimension by Katz.
Source: Authors

4. Fractal characteristic of ECG signal

The calculation of the fractal dimension of Higuchi considers a finite set of observations of time series taken in a regular interval, to then built a new series of time [14]. It is described using Equation (1).

In the work of [18], it is affirmed that the methods based on fractal geometry have been used satisfactorily in the analysis of cardiac signals. These studies show that ECG signals were modeled as self-related fractal sets and their characteristics can be determined using the fractal dimension. In addition, the fractal nature of the ECG signal can be attributed to the self-similar pattern of cardiac rhythm activity, then the change in the rhythmic pattern can be measured using the fractal dimension since the self-similar structure will change.

Likewise, [19], in his article A Healthy Heart is a Fractal Heart, says that the heart is part of a great feedback system, whose dynamics are non-linear, non-stationary and multiscale. Consequently, the heartbeat is one of the most complex signs in nature. Thus, in the records of ECG signals obtained from sick and elderly people, show great losses of complexity with respect to the

beats obtained from healthy people. That is, the values of the fractal index tend to 1 when people are sick and on the contrary they move away from the unit when they are healthy.

Therefore, from a group of 4 ECG signals, the most distorted wave is the signal in Figure B and it is the one that belongs to a healthy person. Then, figure D belongs to the disease called atrial fibrillation, while Figures A and C belong to patients with severe heart failure. They are show in Figure 5.

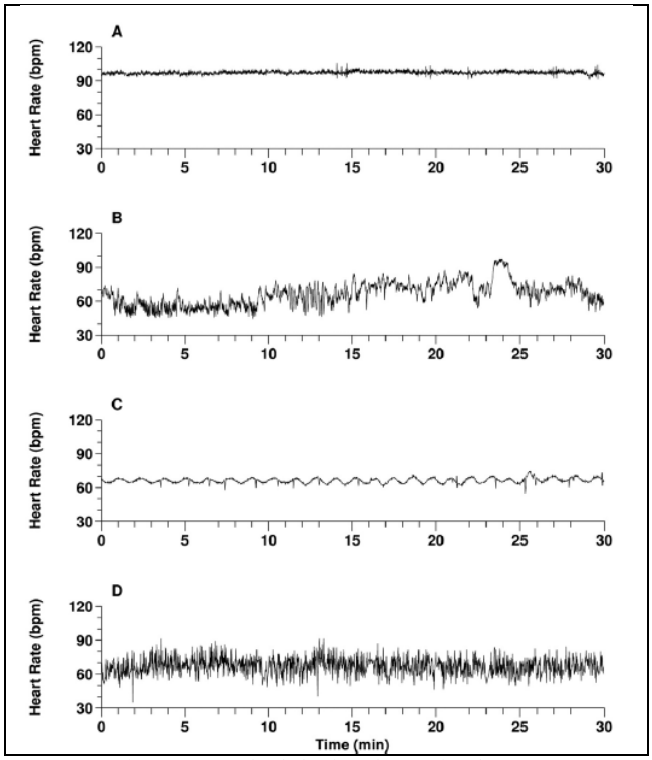


Figure 5. Complex behavior of ECG signals [19].

However, before proceeding to calculate the fractal dimensions of Katz and Higuchi, it is necessary to check their performance. For it, we used the Weierstrass function which was characterized by generating a synthetic signal to verify the operation of the proposed algorithms. Therefore, the procedure used by [20] was repeated, where synthetic signals are generated using the Weierstrass cosine function with known fractal dimensions (theorical), then these signals are processed with the proposed algorithms of Katz and Higuchi. Next, they are represented in tabular and graphic scheme. It is show in Table 1 and in Figure 6.

In the previous graphic representation, was observed a quasi linear behavior of the Higuchi algorithm with respect to the synthetic signal. By another hand, from the graphical representation corresponding to the Katz algorithm, a light variation was observed from the fractal dimension equal to 1.5, showing an increasing behavior as equal as the obtained by [20].

In this way, with the four previously filtered databases, we proceed to calculate the fractal dimensions of Katz and Higuchi in order to differentiate the signals belonging to healthy and sick people. However, such algorithms are of the unsupervised type, because they do not perform learning and they are only used to discriminate or differentiate the data.

Table 1. Averages obtained from the Higuchi and Katz's FD

Source: Authors

FRACTAL DIMENSIONS		
THEORETICAL	**HIGUCHI**	**KATZ**
1.10	1.1298	1.0000
1.20	1.2064	1.0000
1.30	1.294	1.0000
1.40	1.3858	1.0000
1.50	1.4778	1.0000
1.60	1.5700	1.0002
1.70	1.6646	1.0011
1.80	1.7668	1.0059
1.90	1.8797	1.0331
1.95	1.9390	1.0781

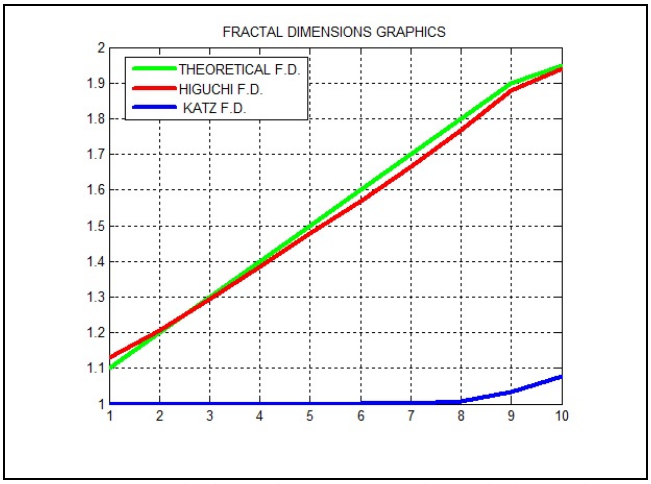


Figure 6. Graph of Fractal Dimensions of the Synthetic Signal. Source: Authors.

For the estimation procedure of both fractal dimensions, each signal of the group of 15 was subdivided into 10 equal sections. For example, for the Physikalisch-Technische Bundesanstalt (PTB) database, belonging to ischemic heart disease, 115,200 samples are collected. However, only the first 100,000 samples were selected and subdivided into 10 equal parts to obtain frames or sections of 10,000 samples with an overlap of 1,000 samples.

KAHF_pura.KA_time_HF_pura <15x10 double>

	1	2	3	4	5	6	7	8	9	10
1	1.2282	1.2121	1.2831	1.2243	1.2199	1.2077	1.2044	1.2125	1.2119	1.2080
2	1.2133	1.2151	1.2182	1.2237	1.2215	1.2225	1.2204	1.2168	1.2240	1.2225
3	1.1854	1.1771	1.1966	1.1935	1.1996	1.1991	1.1880	1.2052	1.1951	1.1953
4	1.1776	1.1482	1.1545	1.1533	1.1825	1.1628	1.1488	1.1527	1.1560	1.1656
5	1.3207	1.3058	1.3514	1.3091	1.3148	1.3314	1.3295	1.3295	1.3207	1.3729
6	1.1808	1.1979	1.2202	1.1875	1.1841	1.1871	1.1956	1.1836	1.1952	1.1912
7	1.2280	1.2215	1.2242	1.2288	1.2279	1.2316	1.2270	1.2270	1.2272	1.2298
8	1.1897	1.1917	1.1815	1.1806	1.1803	1.1792	1.1786	1.1820	1.1913	1.1810
9	1.4409	1.4478	1.4834	1.4435	1.4301	1.4291	1.4560	1.4334	1.4355	1.4345
10	1.2405	1.2265	1.2254	1.2281	1.2259	1.2243	1.2257	1.2261	1.2299	1.2281
11	1.5676	1.5111	1.4853	1.5025	1.5433	1.5290	1.5187	1.5267	1.5272	1.5279
12	1.1783	1.1745	1.1709	1.1706	1.1657	1.1682	1.1768	1.1741	1.1664	1.1700
13	1.2000	1.1980	1.1886	1.1929	1.1904	1.1932	1.1927	1.1937	1.1953	1.1873
14	1.2827	1.2751	1.2737	1.2790	1.2853	1.2578	1.2691	1.3251	1.3278	1.2946
15	1.3413	1.3559	1.3477	1.3968	1.3742	1.4005	1.3690	1.3719	1.3701	1.3727

Figure 7. Katz's FD of Heart Failure.

Figures 7 and 8 show the fractal dimensions of Katz for the 15 signals of Heart Failure and Ischemic Heart Disease, respectively. Similarly, Figures 9 and 10 show the fractal

dimensions of Higuchi for the 15 signals of Normal Signal and Signal with Hypertension, respectively.

KAMI_pura.KA_time_MI_pura <15x10 double>

	1	2	3	4	5	6	7	8	9	10
1	1.4576	1.4507	1.4464	1.4504	1.4337	1.4412	1.4316	1.4365	1.4348	1.4317
2	1.4422	1.4747	1.4814	1.4679	1.4421	1.4468	1.4627	1.4387	1.4422	1.4401
3	1.4088	1.4058	1.4038	1.4267	1.4281	1.3944	1.3957	1.3890	1.4008	1.4104
4	1.3606	1.3699	1.3558	1.3536	1.3449	1.3226	1.3215	1.3276	1.3280	1.3299
5	1.3326	1.3421	1.3388	1.3460	1.3393	1.3461	1.3394	1.3369	1.3372	1.3392
6	1.4761	1.4973	1.4852	1.4969	1.4906	1.4995	1.4918	1.4863	1.4916	1.4879
7	1.3718	1.3788	1.3643	1.3719	1.3712	1.3695	1.3778	1.3741	1.3646	1.3727
8	1.3605	1.3528	1.3483	1.3515	1.3570	1.3668	1.3615	1.3482	1.3536	1.3514
9	1.3863	1.3790	1.3856	1.3838	1.3820	1.3880	1.4021	1.3957	1.3919	1.3872
10	1.3636	1.3650	1.3650	1.3980	1.4249	1.4248	1.4312	1.4202	1.4186	1.4111
11	1.3260	1.3508	1.3355	1.3771	1.3760	1.3358	1.3144	1.3359	1.2873	1.2655
12	1.3454	1.3490	1.3458	1.3442	1.3465	1.3456	1.3456	1.3477	1.3473	1.3494
13	1.4305	1.3815	1.3805	1.4146	1.4351	1.4510	1.4172	1.3932	1.3826	1.3817
14	1.4948	1.4740	1.4591	1.4664	1.4620	1.4593	1.4710	1.4696	1.4642	1.4642
15	1.3968	1.3979	1.4043	1.4064	1.4079	1.4059	1.4086	1.4081	1.4088	1.4111

Figure 8. Katz's FD of Ischemic Heart Disease.

FD_time_SR_pura <15x10 double>

	1	2	3	4	5	6	7	8	9	10
1	1.8313	1.8359	1.8345	1.8286	1.8444	1.8463	1.8435	1.8603	1.8409	1.8132
2	1.7720	1.7871	1.8289	1.8300	1.8484	1.8733	1.8945	1.8462	1.8558	1.8321
3	1.7943	1.7957	1.8041	1.8217	1.8116	1.8202	1.8165	1.8065	1.8101	1.7888
4	1.8741	1.8885	1.8449	1.8641	1.8879	1.8863	1.8858	1.8681	1.8796	1.8810
5	1.8139	1.7984	1.8001	1.8037	1.7919	1.7785	1.7648	1.7677	1.7728	1.7866
6	1.7859	1.7810	1.7947	1.8057	1.7905	1.8134	1.8052	1.8134	1.8375	1.7965
7	1.7769	1.7612	1.7808	1.7527	1.7905	1.8113	1.7693	1.7724	1.7576	1.7505
8	1.8041	1.8097	1.8236	1.8245	1.8160	1.8117	1.8061	1.8209	1.8137	1.8171
9	1.7249	1.7494	1.7321	1.7466	1.7285	1.7124	1.7010	1.7356	1.8162	1.7634
10	1.8153	1.8442	1.8241	1.8290	1.8231	1.8243	1.8385	1.8389	1.8222	1.8376
11	1.7965	1.8273	1.8291	1.8296	1.8172	1.8429	1.8123	1.8316	1.8542	1.8331
12	1.8152	1.8228	1.8277	1.8410	1.8300	1.8205	1.8236	1.8331	1.8166	1.8169
13	1.8279	1.8237	1.8808	1.8455	1.8632	1.8330	1.8192	1.8311	1.8599	1.8800
14	1.4197	1.7962	1.7916	1.8081	1.7882	1.7443	1.7316	1.7200	1.7497	1.7971
15	1.5084	1.6679	1.6655	1.6696	1.6735	1.6755	1.6748	1.6741	1.6737	1.6816

Figure 9. Higuchi's FD of the Normal Sinus Rhythm.

FD_time_SP_pura <15x10 double>

	1	2	3	4	5	6	7	8	9	10
1	1.6237	1.6878	1.6477	1.5958	1.5899	1.6393	1.5768	1.6427	1.6099	1.5667
2	1.6729	1.6764	1.6640	1.6629	1.6775	1.6729	1.6706	1.6747	1.6747	1.6880
3	1.6696	1.6709	1.7682	1.7087	1.6639	1.6582	1.6615	1.6452	1.7047	1.7189
4	1.6929	1.7024	1.7038	1.7181	1.7113	1.6833	1.6783	1.6970	1.6957	1.6964
5	1.7368	1.8189	1.7871	1.7051	1.6840	1.6711	1.6893	1.6791	1.6725	1.7053
6	1.4290	1.4685	1.5152	1.6608	1.6358	1.6846	1.6382	1.6306	1.6724	1.5931
7	1.6442	1.6700	1.6272	1.5432	1.5198	1.6527	1.6789	1.6744	1.6339	1.6533
8	1.6055	1.5776	1.5977	1.6124	1.5940	1.6108	1.5961	1.5395	1.5971	1.5830
9	1.4463	1.4669	1.4772	1.4794	1.4874	1.4843	1.4724	1.4859	1.4879	1.4957
10	1.4664	1.4787	1.4639	1.4733	1.4759	1.5348	1.4898	1.5110	1.5058	1.4668
11	1.5187	1.5158	1.5907	1.5332	1.5299	1.5378	1.5181	1.5604	1.5596	1.6246
12	1.6558	1.7587	1.6853	1.7012	1.6913	1.6682	1.6713	1.6773	1.6893	1.6726
13	1.5948	1.6298	1.6056	1.6320	1.6124	1.6342	1.6571	1.6520	1.6636	1.6490
14	1.4415	1.4447	1.4528	1.4569	1.4361	1.4581	1.4495	1.4692	1.4500	1.4379
15	1.4171	1.4187	1.4092	1.4291	1.4184	1.4321	1.4230	1.4241	1.4249	1.4236

Figure 10. Higuchi's FD of the Signal with Hypertension.

As observed from Figures 7, 8, 9 and 10, there are 10 dimensions for each of the 15 signals in the database. Therefore, we proceeded to determine the arithmetic mean of the 10 FD, because are very close to each other. They are show in Tables 2 and 3.

However, with the arithmetic mean it is not possible to discriminate easily between one and another type of ECG signal, because such result present not scattered values for the same type of ECG signal. Only in the case of normal signal and ischemic heart disease do not have so scattered values. Otherwise, in the case of the Higuchi algorithm, the results of the arithmetic means present closer values, and they are no higher as scattered for each of the four signals, especially in the case of the normal sinus rhythm (healthy people).

Table 2. Averages obtained from the Katz's FD

Heart Failure	Ischemic Heart Disease	Normal Sinus Rhythm	Signal with Hypertension
1.2212	1.4415	1.5761	1.3590
1.2198	1.4539	1.4090	1.3538
1.1935	1.4064	1.5360	1.2828
1.1602	1.3415	1.4395	1.2777
1.3286	1.3398	1.4750	1.3631
1.1923	1.4903	1.4491	1.7435
1.2273	1.3717	1.5331	1.8564
1.1836	1.3552	1.4787	1.5952
1.4434	1.3882	1.3967	1.6198
1.2280	1.4022	1.3928	1.5117
1.5239	1.3304	1.4818	1.5321
1.1715	1.3466	1.4576	1.6883
1.1932	1.4068	1.5216	1.5666
1.2870	1.4685	1.3166	1.7599
1.3700	1.4056	1.4618	1.7354

Table 3. Averages obtained from the Higuchi's FD

Heart Failure	Ischemic Heart Disease	Normal Sinus Rhythm	Signal with Hypertension
1.3629	1.5116	1.8379	1.6180
1.6130	1.5127	1.8368	1.6735
1.3751	1.4801	1.8069	1.6870
1.6722	1.3048	1.8760	1.6979
1.5862	1.4361	1.7878	1.7149
1.5137	1.4550	1.8024	1.5928
1.5747	1.4296	1.7723	1.6298
1.6073	1.4601	1.8147	1.5914
1.5717	1.4552	1.7410	1.4783
1.5602	1.3781	1.8297	1.4866
1.4120	1.4407	1.8274	1.5489
1.4640	1.4019	1.8247	1.6871
1.6003	1.4811	1.8464	1.6330
1.4542	1.5284	1.7346	1.4497
1.5090	1.2369	1.6565	1.4220

In this way, in order to have a better visualization and differentiation of the obtained results, we processed the FD by calculating the variance in order to see how these results differ from their average. Results of the variance of the averages obtained from the FD of Higuchi and Katz, for the four types of signals used, are shown. Also, signal with the greatest variation for the Higuchi's FD is arterial hypertension and the one with the least variation is the FD of the normal sinus rhythm. Similarly, for the Katz algorithm, the FD with the greatest variation is also arterial hypertension and the one with the least variation is that of ischemic heart disease. It is show in Table 4.

Then, in order to reduce the dimensionality, we decided to use the Principal Component Analysis on the obtained data. So, in this way, we achieved a better differentiation of the databases studied. For this, we created a matrix where the values of the FD are analyzed.

Table 4. Variance of the averages obtained from the Higuchi and Katz's FD

CARDIOVASCULAR DISEASES	HIGUCHI's F.D.	KATZ's F.D.
Heart Failure	0.0087	0.0116
Normal Signal	0.0030	0.0044
Arterial Hypertension	0.0093	0.0352
Ischemic heart disease	0.0062	0.0025

When the results of the PCA are plotted, the values of the "score" matrix are used, which is generated when working with the PRINCOMP function of the Matlab. This matrix score contains the coordinates or projections of the main components. In a first stage, we used the PCA to compare the signals in pairs, so, the FD of the signals corresponding to the diseases had compared with the signal of the healthy person. In this way, we show some results in two dimensions that comes of the application of the PCA algorithm on the FD of Katz and Higuchi, respectively, for the case of frame equal to 10,000 and overlap equal to 1,000.

Thus, for Katz's Fractal Dimension, the comparison of heart failure with the normal signal showed a tendency between the two when trying to group, but several scattered points were also observed. It is show in Figure 11.

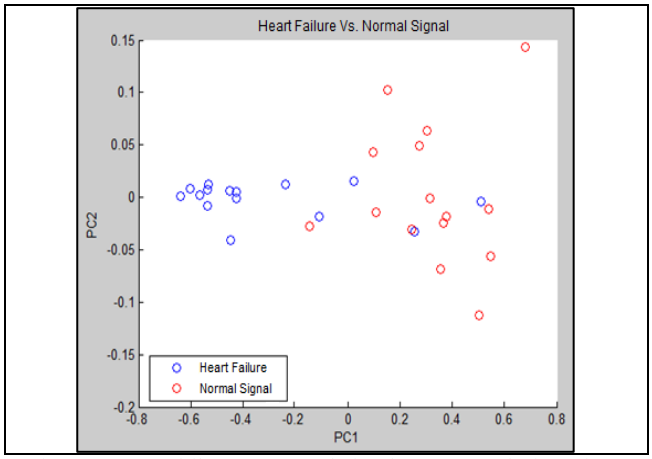


Figure 11. Heart failure Vs. Normal Signal (Katz Fractal Dimension). Frame=10,000 and Overlap=1,000.

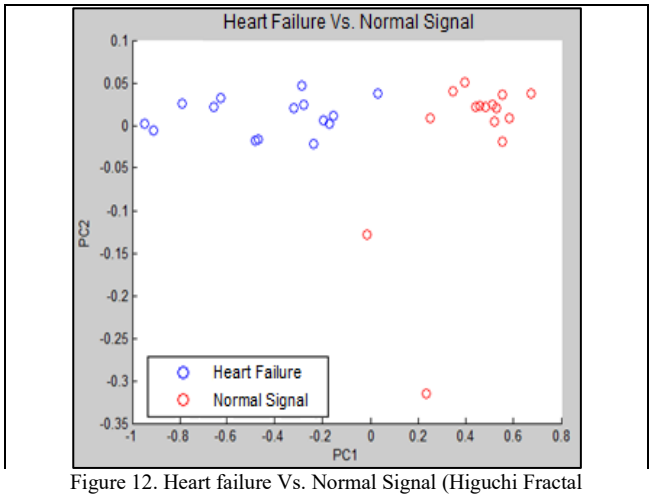


Figure 12. Heart failure Vs. Normal Signal (Higuchi Fractal Dimension). Frame=10,000 and Overlap=1,000.

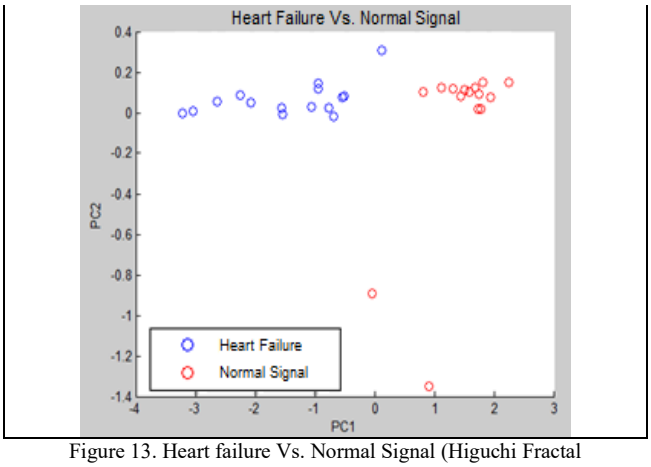


Figure 13. Heart failure Vs. Normal Signal (Higuchi Fractal Dimension). Frame=1,000 and Overlap=100.

On the other hand, for Higuchi's Fractal Dimension, the comparison of heart failure with the normal signal showed a better separation between the 2 signals observed. It is show in Figure 12.

Performing more tests with the algorithms used, we proceeded to reduce the size of the frame to 1,000 and the overlap to 100. This gave rise to a calculation of the FD for 100 segments and therefore more features emerged were analyzed with the PCA algorithm. For this case, again, the Higuchi algorithm presented a better performance to analyze and differentiate the signals.

Similarly, for Higuchi's Fractal Dimension, the comparison of heart failure with the normal signal showed an excellent differentiation of the main components. It is show in Figure 13.

In this way, for Katz's Fractal Dimension, the comparison of heart failure with the normal signal showed an improvement in the differentiation of the main components. It is shown in Figure 14.

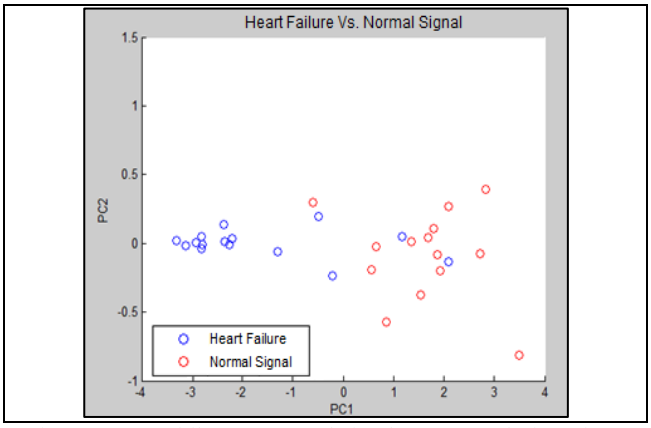


Figure 14. Heart failure Vs. Normal Signal (Katz Fractal Dimension). Frame=1,000 and Overlap=100.

Therefore, with the reduction of the size of the frame and of the overlap a major separation of the main components is achieved.

At the end, we show, in 2D and 3D graphs, principal components (PCA) of the 4 signals studied for Higuchi Fractal Dimension for the case of frame equal to 1,000 and over equal to 100, where it is appreciated that heart failure is mixed with hypertension and ischemic heart disease. They are show in Figures 15 and 16.

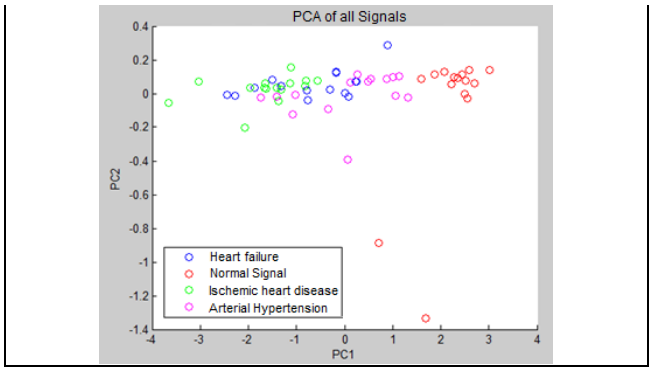


Figure 15. 2D graph of the resulting PCA (Higuchi Fractal Dimension). Frame=1,000 and Overlap=100.

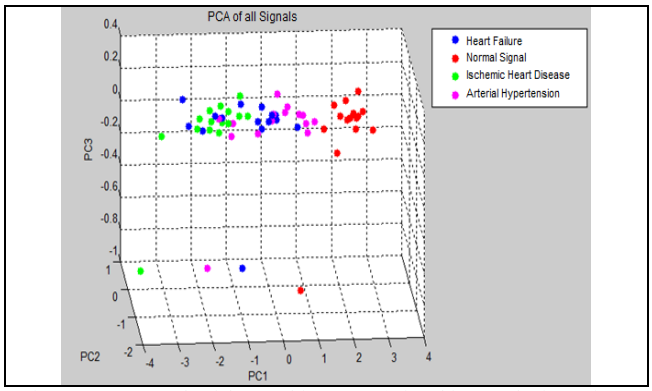


Figure 16. 3D graph of the resulting PCA (Higuchi Fractal Dimension). Frame=1,000 and Overlap=100.

On the other hand, as a result of the calculation of Katz´s Fractal Dimension for the case of frame equal to 1,000 and over equal to 100, the graphic representation in 2D and 3D of the main components of te 4 signals studied show a greater combination between them, making their separation and identification more difficult. They are show in Figures 17 and 18.

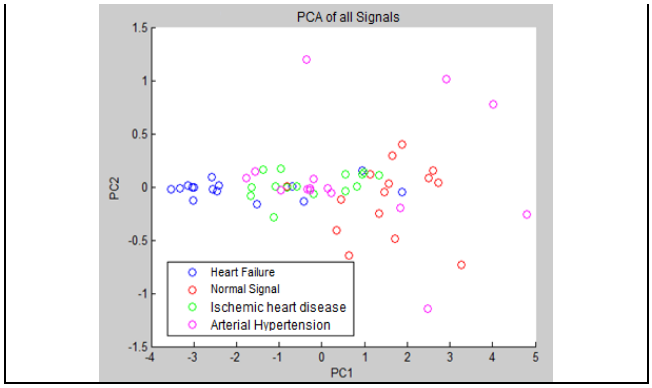


Figure 17. 2D graph of the resulting PCA (Katz Fractal Dimension). Frame=1,000 and Overlap=100.

5. Conclusions

The Butterworth high pass filter was used for the elimination of the variations of the baseline, which allowed to filter the frequencies below 0.5 Hz, range in which they are present unwanted signals. In addition, the algorithms of Katz and Higuchi were adapted to the programming language of Matlab, for this purpose the corresponding equations were used, which were the basis of calculation for these algorithms.

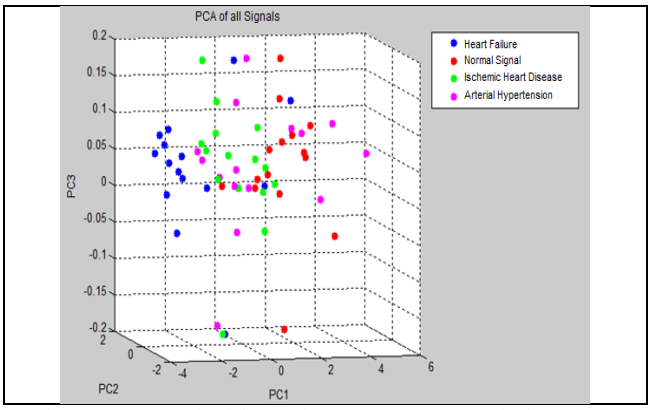


Figure 18. 3D graph of the resulting PCA (Katz Fractal Dimension). Frame=1,000 and Overlap=100.

PCA algorithm was used optimally, since it was possible to reduce the FD, obtaining the main components that represent the 4 types of signals (heart failure, hypertension, ischemic heart disease and normal signal).

The results were graphically represented with the principal components in two and three dimensions for the Katz and Higuchi algorithms. Finally, the algorithm of Higuchi gave a better performance and this is because Higuchi's algorithm calculates the FD through summations and repetitive averages of the segment or distance at which the FD was calculated, while Katz uses the basic form of the calculation proposed by Mandelbrot when define a fractal, which consists of a division of logarithms to calculate the FD. This performance results when refer to the observation of the variance of the FD averages, which are shown in Table 4.

In addition, the results improve when the size of the frame and the over are reduced to 1,000 and 100, respectively. This improvement is due to the fact that the components of the PCA do not appear dispersed, and on the contrary they remain grouped which facilitates their differentiation. In this way, despite using a frame size equal to 10,000 and an overlap equal to 1,000, the Higuchi algorithm achieved a better separation of the PCA components from the databases: normal signal y heart failure, as show in figures 11 and 12. For the other hand, by reducing the number of frames to 1,000 and the overlap to 100, again the Higuchi algorithm achieves a better grouping of the components, as show in figures 16 and 18.

On the other hand, the result of the work of [20] shows that Katz's algorithm is the most consistent method for discrimination of epileptic states from the intracranial EEG (IEEG), likely due to its exponential transformation of FD values and relative insensitivity to noise. For this reason, the content of this article differs in terms of the results achieved with respect to Higuchi's algorithm.

Finally, it can also be concluded that the use of fractal dimensions plus other techniques such as artificial neural networks, frequency analysis, principals components analysis, among others, offer greater applications of identification and classification on ECG and EEG signals allowing the early detection of Alzheimer's, identifying cardiovascular diseases, predicting sudden deaths due to heart problems, classifying heart signals [3], [7], [11], [12], etc.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] World Health Organization. Press Center. Cardiovascular diseases. 2015. Recovered in June, 2017. Available: http://www.who.int/mediacentre/factsheets/fs317/es/

[2] Physiobank Database, 2017. Available: https://physionet.org/physiobank/database/

[3] A. Al-nuaimi, E. Jammeh, L. Sun and E. Ifeachor, "Higuchi fractal dimension of the electroencephalogram as a biomarker for early detection of Alzheimer's disease". In 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2017. DOI: 10.1109/EMBC.2017.8037320.

[4] S. Solehah, V. Sagayan, S. Chandra and D. Yanti, "Evaluation of simulated VEP signals on basis of Higuchi and Katz's algorithm". In 2017 IEEE International Conference on Signal and Image Processing Applications, 2017. DOI: 10.1109/ICSIPA.2017.8120627.

[5] C. Flores and J. Noel, "Parameters analyzed of Higuchi's fractal dimension for EEG brain signals". In 2015 Signal Processing Symposium (SPSympo), Debe, Poland, 2015. DOI: 10.1109/SPS.2015.7168285.

[6] S. Spasić, A. Savic, L. Nikolic, S. Budimir, D. Janošević and A. Mitrović, "Applications of Higuchi's fractal dimension in the analysis of biological signals". In 2012 20th Telecommunications Forum (TELFOR), 2012. DOI: 10.1109/TELFOR.2012.6419291.

[7] I. Sedielmaci & F. Bereksi, "Detection of some heart diseases using fractal dimension and chaos theory". In 2013 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA), 2013. DOI: 10.1109/WoSSPA.2013.6602342

[8] A. Gómez, O. López, D. Sierra y O. Rueda-Ochoa, "Extracción de características fractales de la señal de variabilidad de frecuencia cardiaca en pacientes con accidente cerebrovascular y análisis de su viabilidad como factor pronóstico". In XVI Simposio de Tratamiento de Señales, Imágenes y Visión Artificial (STSIVA), 2011.

[9] K. Amit and S. Raghav, "Local fractal dimension based ECG arrhythmia classification" Journal Biomedical Signal Processing and Control, 5(2), 114–123, 2010. https://doi.org/10.1016/j.bspc.2010.01.002

[10] U. Rajendra, H. Fujita, M. Adam, O. Shu, T. Jen, V. Sudarshan and J. EW Koh, "Automated characterization of arrhythmias using nonlinear features from tachycardia ECG beats". In 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary. 2016. DOI: 10.1109/SMC.2016.7844294.

[11] F. López-Caracheo, A. Bazualda, C. Perez-Ramirez, M. Valtierra-Rodriguez, A. Dominguez-Gonzalez, and J. Amezquita-Sanchez, "Fractal Dimension-based Methodology for Sudden Cardiac Death Prediction". In IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC 2018), Ixtapa, Mexico. 2018. DOI: 10.1109/ropec.2018.8661371.

[12] M. Hamidi, H. Ghassemian and M. Imani, "Classification of heart sound signal using curve fitting and fractal dimension" Journal Biomedical Signal Processing and Control, 39, 351-359, 2018. https://doi.org/10.1016/j.bspc.2017.08.002

[13] Recommendations for the Standardization and Interpretation of the Electrocardiogram: Part I. 2007 Available: https://www.sciencedirect.com/science/article/pii/S073510970700232X#bib23

[14] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory" Journal Physica D: Nonlinear Phenomena, 31(2), 277-283, 1988. https://doi.org/10.1016/0167-2789(88)90081-4

[15] S. Selvam, "Complete Higuchi Fractal Dimension Algorithm". Mathworks, File Exchange. Recovered in may, 2017. Available: https://la.mathworks.com/matlabcentral/fileexchange/30119-complete-higuchi-fractal-dimension-algorithm

[16] M. Katz, "Fractals and the Analysis of Waveforms" Journal Computers in Biology and Medicine, 18(3), 145-156, 1988. https://doi.org/10.1016/0010-4825(88)90041-8

[17] J. Monge-Alvarez, "Higuchi and Katz fractal dimension measures". Mathworks, File Exchange. Recovered in may, 2017. Available: https://la.mathworks.com/matlabcentral/fileexchange/50290-higuchi-and-katz-fractal-dimension-measures

[18] D. Kumar, Fractals Applications in Biological Signaling and Image Processing. Boca Raton, USA: Taylor & Francis Group, 2017.

[19] Cipra B, "A healthy Heart is a Fractal Heart. SIAM News", 2003. Available: https://studylib.net/doc/10393710/a-healthy-heart-is-a-fractal-heart-by-barry-a.-cipra

[20] R. Esteller, G. Vachtsevanos, J. Echauz and B. Litt, "A comparison of waveform fractal dimension algorithms". In IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications, 2001. DOI: 10.1109/81.904882.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 151-157 (2019)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Satisfaction of the Graduate for the Continuous Improvement of Educational Quality in UNTELS

Omar Freddy Chamorro Atalaya[1*], Teodoro Neri Díaz Leyva[1], Dora Yvonne Arce Santillan[1], Jorge Isaac Castro Bedriñana[1], Denisse Marie Barrientos Pichilingue[1], Avid Roman-Gonzalez[2]

[1]*Faculty of Engineering and Management, Universidad Nacional Tecnológica de Lima Sur, Lima, Peru*

[2]*Aerospace Sciences and Health Research Laboratory (INCAS-Lab), Universidad Nacional Tecnológica de Lima Sur, Lima, Peru*

ARTICLE INFO

Article history:
Received: 12 July, 2019
Accepted: 01 September, 2019
Online: 24 September, 2019

Keywords:
Student satisfaction
Graduated
Continuous Improvement
Competitions
Personal attitudes
Teaching plan
Library services
Laboratory services
Computer center
Administrative service
Support service

ABSTRACT

The Universidad Nacional Tecnologica de Lima Sur (UNTELS), like all universities, has as one of its objectives the training of professionals with skills that the changing working world requires. One way to establish the relevance of what UNTELS offers to its students is to know relevant information from graduates that entails, to their authorities, establish policies for continuous improvement of educational quality and improve the training of new professionals. In this context, the purpose of this research project is to collect information from graduates of the Professional Careers of Systems Engineering, Environmental Engineering, Mechanical and Electrical Engineering, Electronic Engineering and Communications, Business Administration from the National Technological University of South Lima of the second academic semester of the year 2017; To evaluate the satisfaction with regard to their professional training, receive a physical questionnaire with questions aimed at knowing their professional competencies, teaching staff, environments, library services, laboratory services, computer center services, administrative service of the Faculty and Professional School, University support services and personal and social attitudes that support decision-making for the improvement of educational quality at the National Technological University of South Lima.

1. Introduction

The satisfaction of the educational quality of students or graduates of higher education institutions (IES) becomes more relevant because of the social relevance of higher education and the competition between the different study options of the most affluent or most talented students [1]

In a study of student satisfaction of the Ana G. Méndez University System, it tells us that the average level of overall satisfaction towards university services worldwide is 78% [2]

Likewise, according to Llanes Castillo and collaborators, it was evident that the satisfaction of the graduates in a Mexican university in terms of the training received, ranged between 54 and 65%, was evaluated regularly by 17% and good by 58.5%. Another similar study carried out in a medical university in Colombia, when exploring the aspects to improve in training, found teaching deficiencies (7.0%), deficiencies in practical activities (19.5%), deficit of texts in the library (0,9%). While a study of graduates in Peru, I detected 6.2% dissatisfaction, and 59.1% reported satisfaction, so that more than 80% of respondents rated between 4 and 5 the general training received in relation to performance as professionals [3].

Satisfaction with the training received is a subjective concept, since it depends on the fulfillment (or not) of the expectations and wishes of the students [4] In addition, it is a broad concept, whose study is conditioned by a great quantity of characteristics, as well as the existence of different points of view to address its measurement [5].

Among the latter, there is the vision provided by Lapeña and González (1996), which indicate that the degree of satisfaction of a student with the training received is related to their opinion on it (which highlights the subjectivized commented) [6]. According to these authors, it would be an attempt to see to what extent the

* Omar Freddy Chamorro Atalaya, Virgen de la Puerta, Los Damascos 986, Los Olivos, Lima, Perú, 968053089 & omar_chamorro1@hotmail.com

student considers training in the performance of their work or training useful and if they liked it [7].

Thus, the continuous improvement of university quality is not only due to increasing resources and infrastructures in academic units, but mainly by launching trained professionals with optimal academic performance and, above all, satisfied with the preparation received [8].

In this sense, the National Technological University of Lima Sur (/UNTELS) aims to have information on the satisfaction of graduates regarding their professional training, so that the managers of the National Technological University of Lima South support the taking of decisions to adapt their professional training of students to the existing demand of the productive environments.

The present work is structured as follows. Section 2 presents the methodology used in this research work; this methodology will be descriptive / correlational, with which the relationship of the indicators that present the highest level of dissatisfaction will be found. Section 3 shows the results obtained after the surveys carried out. In Section 4 he presents the statistical analysis. Finally, in Section 4 the discussion of the results obtained is presented.

2. Methodology

2.1. Kind of investigation

The research is basic; the level is descriptive/ correlational.

2.2. Population and Sample

The population is made up of the 160 graduates of the second academic semester of 2017 of the professional careers of Systems Engineering, Mechanical and Electrical Engineering, Environmental Engineering, Business Administration and Electronic and Telecommunications Engineering of the National Technological University of Lima Sur.

The sample is equal to the population and is not probabilistic, because it was only considered, to apply the questionnaire to graduates of the second academic semester of 2017, which corresponds to a number of 160 graduates; these results were obtained as part of the application of a questionnaire made to all graduates. It should be noted that UNTELS University is an institution with few years of operation specifically founded in 2007, and its first graduate promotion is registered in 2011, to date the University has no report or reference base or point Starting that allows taking actions in relation to the satisfaction of the graduates.

Then the type of non-probabilistic sampling that was used is "deliberate, critical or trial sampling", which is a technique in which the members of the sample are chosen only on the basis of the knowledge and judgment of the investigator. As the knowledge of the researcher is instrumental in creating a sample, there are chances that the results obtained are highly accurate with a minimum margin of error. This type of sampling is more effective in situations where there are only a limited number of people they possess the same qualities that a researcher expects from the target population [9].

There is no bias in the data because the questionnaire was applied to the entire population formed by the 160 graduates of all professional careers, without considering any distinction; the self-selection was possible since the presence of the 160 graduates was counted, thus having a form of control over the possibility of their participation. Due to this control it was possible to obtain all the answers requested in the questionnaire.

2.3. Applied Techniques in the Collection of Informacion and Measuring Instruments

A questionnaire with opinion questions about the satisfaction of the graduate of the quality of the different services offered by UNTELS was used as an instrument, considering indicators regarding knowing their professional competencies, teaching staff, environments, library services, laboratory services, services of computer center, administrative service of the Faculty and Professional School, support services of the University and personal and social attitudes; Because it is the first time this type of research is carried out, it was decided to opt for a simple, easy to understand and low cost tool for the collection of information.

Questionnaires were used to measure satisfaction, since according to Franklin and Walker (2010), the use of questionnaires has the advantages of speed, moderate cost, and can cover a wide variety of potential problems ranging from spelling errors, grammar or of writing up problems with the concepts that are being operationalized in the questionnaire, additionally covers cognitive aspects of the respondents, can reveal potential difficulties for the interviewer (if there is one) and can reveal possible problems for data analysis. (Landa, M. & Ramírez, M., 2017). [9]

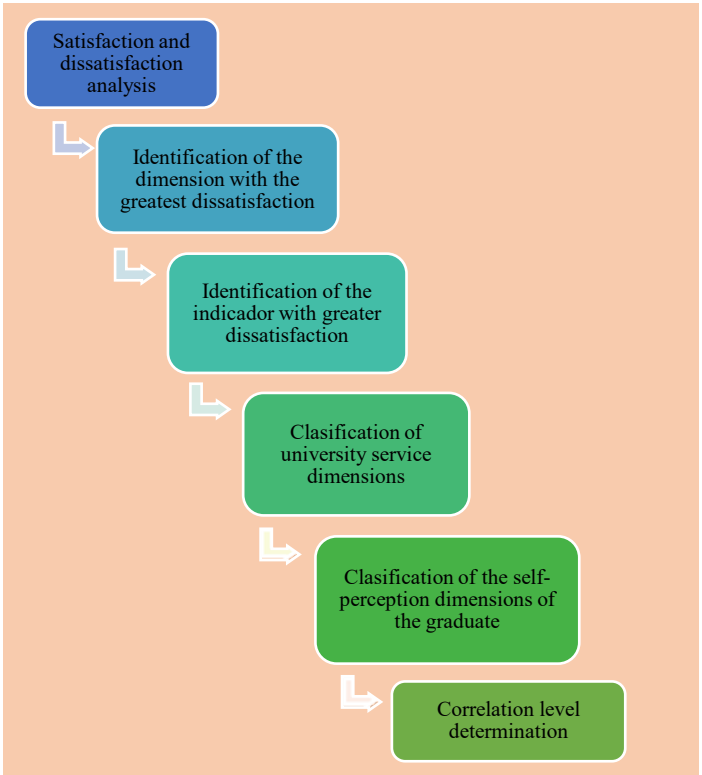


Figure 1: Sequence of the investigation procedure

There are other methods to measure student satisfaction, such is the case of the SERVQUAL method, but according to the empirical studies carried out by Cronin and Taylor, this is not the most appropriate evaluations, due to the frequent interpretation problems that it poses to the people to whom questions are administered; They also represent a redundancy within the measuring instrument, since perceptions are influenced by expectations, while the questionnaires are one of the most economical ways of obtaining quantitative data and a wide scope; They also allow you to analyze the results easily with integrated tools, without having to have knowledge of statistics or scientific research.

The following figure shows the sequence of the procedure that the present investigation has.

3. Results

The results obtained from the application of the data collection instrument of the "Graduate Satisfaction Survey", for the second academic semester of 2017, are shown below.

The following Figure 2 shows the results of the Indicators with the highest levels of dissatisfaction by specialty of the Professional Competencies dimension.

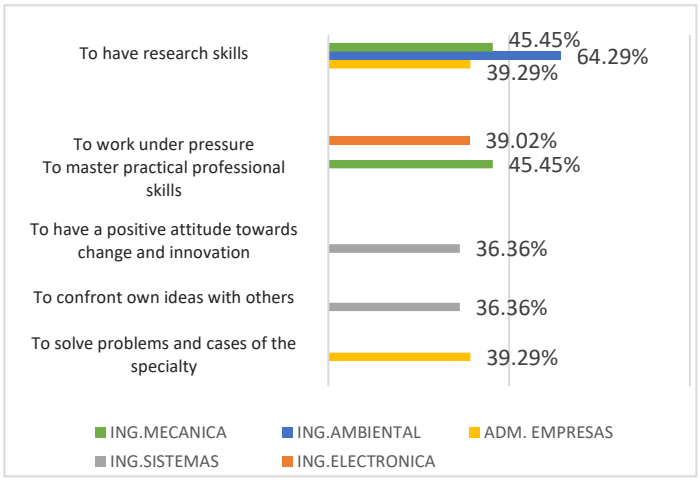


Figure 2: Indicators with the highest dissatisfaction levels by specialty of the dimension professional competences

As can be seen in Figure 2, the indicator that shows the highest percentage of dissatisfaction is "To have research skills" equivalent to 64.29% in the specialty of Environmental Engineering in the Professional Competencies dimension.

The following Figure 3 shows the results of the Indicators with the highest levels of dissatisfaction by specialty of the Teaching Plan dimension.

As can be seen in Figure 3, the indicator that shows the highest percentage of dissatisfaction is "Proficiency in the subjects of the courses it develops" equivalent to 64.29% in the specialty of Environmental Engineering in the Teaching Plane dimension.

The following Figure 4 shows the results of the Indicators with the highest levels of dissatisfaction by specialty of the University Environments dimension.

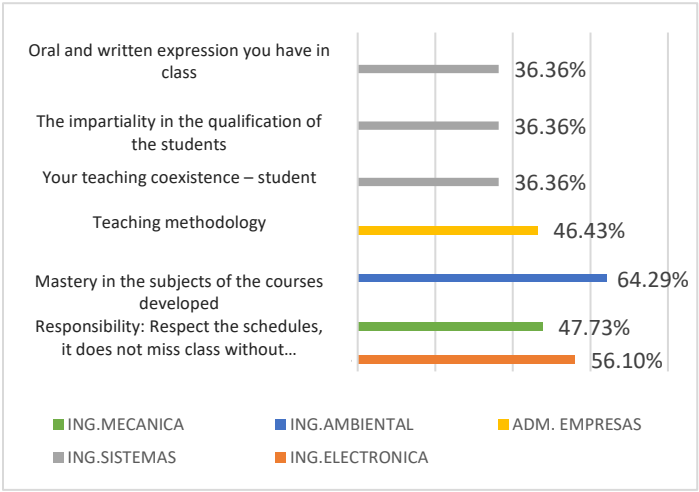


Figure 3: Indicators with the highest dissatisfaction levels by specialty of the teaching flat dimension

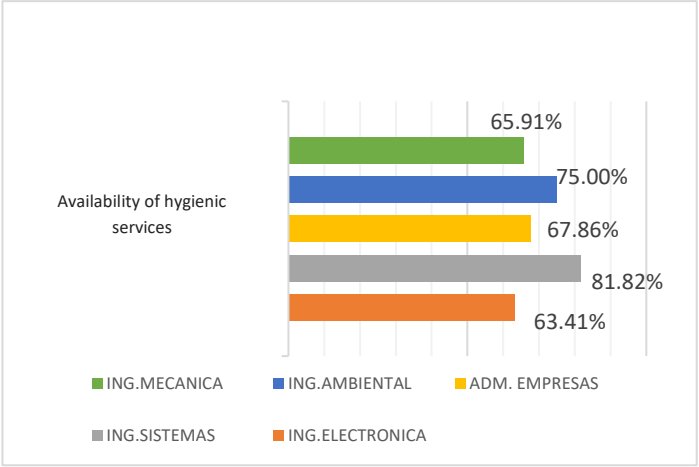


Figure 4: Indicators with the highest dissatisfaction levels by specialty of the dimension environments of the University

As shown in Figure 4, the indicator that shows the highest percentage of dissatisfaction is "Availability of Hygienic Services" equivalent to 81.82% in the specialty of Systems Engineering in the University Environments dimension.

The following Figure 5 shows the results of the Indicators with the highest levels of dissatisfaction by specialty of the Library Services dimension.

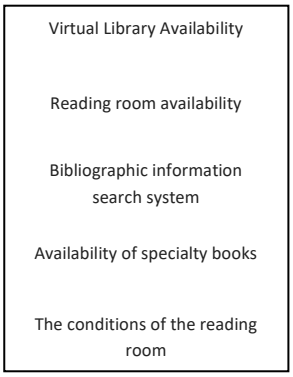


Figure 5: Indicators with the highest dissatisfaction levels by specialty of the dimension library services

As shown in Figure 5, the indicator that shows the highest percentage of dissatisfaction is "Virtual library availability" equivalent to 61.11% in the specialty of Environmental Engineering in the Library Services dimension

The following Figure 6 shows the results of the Indicators with the highest levels of dissatisfaction by specialty of the Laboratory Services dimension.

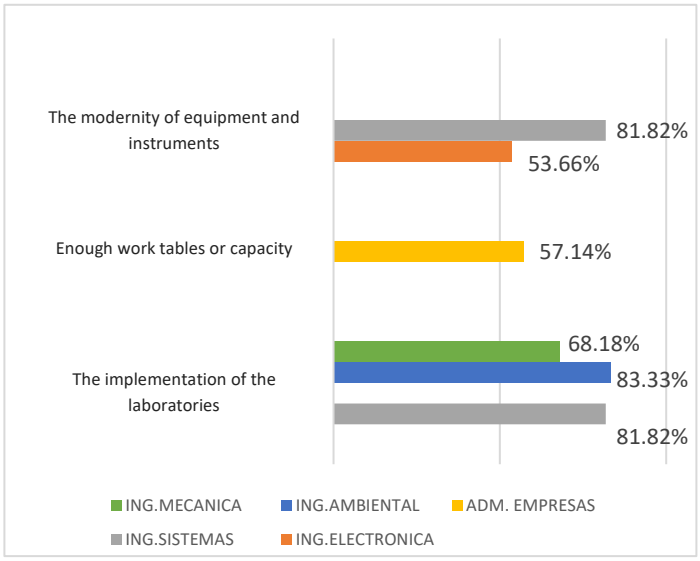


Figure 6: Indicators with the highest dissatisfaction levels by dimension specialty laboratory services

As shown in Figure 6, the indicator that shows the highest percentage of dissatisfaction is "The implementation of laboratories" equivalent to 83.33% in the specialty of Environmental Engineering in the Laboratory Services dimension.

The following Figure 7 shows the results of the Indicators with the highest levels of dissatisfaction by specialty of the Computer Center dimension.

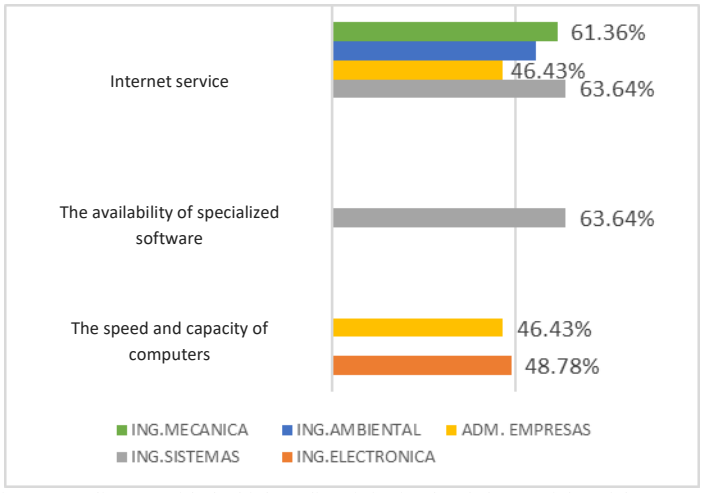


Figure 7: Indicators with the highest dissatisfaction levels by specialty of the dimension computer center

As shown in Figure 7, the indicators that show the highest percentage of dissatisfaction are "Internet service and availability of specialized Software" equivalent to 63.64% in the specialty of Systems Engineering in the Computer Center dimension.

The following Figure 8 shows the results of the Indicators with the highest levels of dissatisfaction by specialty of the Administrative Service dimension of the Faculty and Professional School.

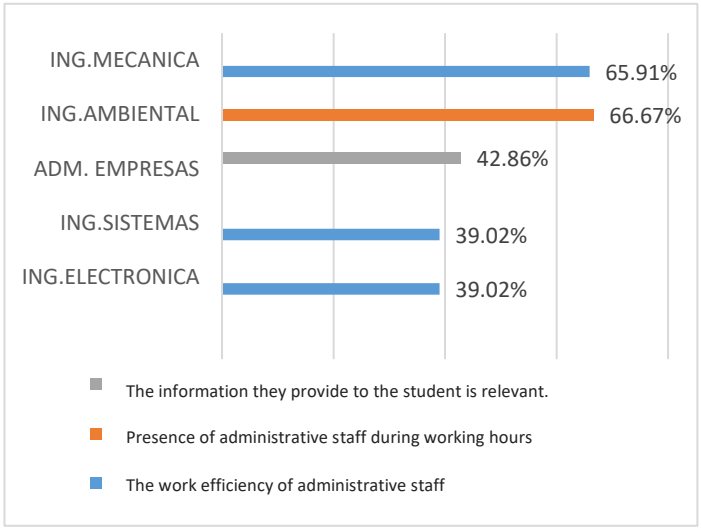


Figure 8: Indicators with the highest insatisfaction levels by specialty of the dimension administrative service of the school and professional school

As shown in Figure 8, the indicator that shows the highest percentage of dissatisfaction is "Presence of administrative staff during working hours" equivalent to 66.67% in the specialty of Environmental Engineering in the Administrative Service dimension of the Faculty and Professional School.

The following Figure 9 shows the results of the Indicators with the highest levels of dissatisfaction by specialty of the Support Services dimension of the University.

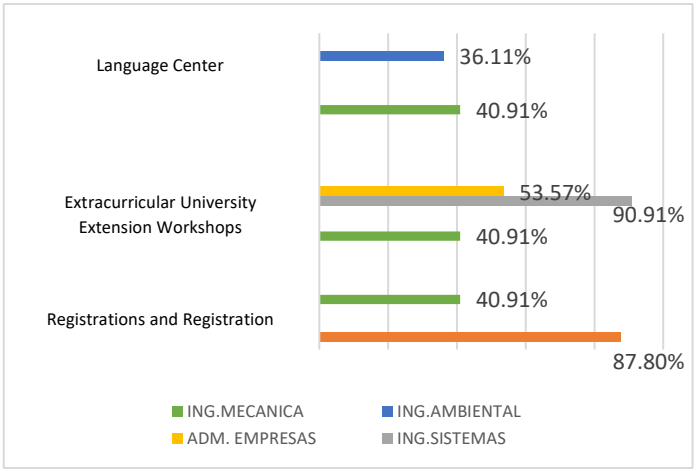


Figure 9: Indicators with the highest dissatisfaction levels by specialty of the dimension support services of the University

As can be seen in Figure 9, the indicator that shows the highest percentage of dissatisfaction is "Registries and enrollment" equivalent to 90.91% in the specialty of Systems Engineering in the Support Services dimension of the University.

The following Figure 10 shows the results of the Indicators with the highest levels of dissatisfaction by specialty of the Personal and Social Attitudes (Self-perception) dimension.

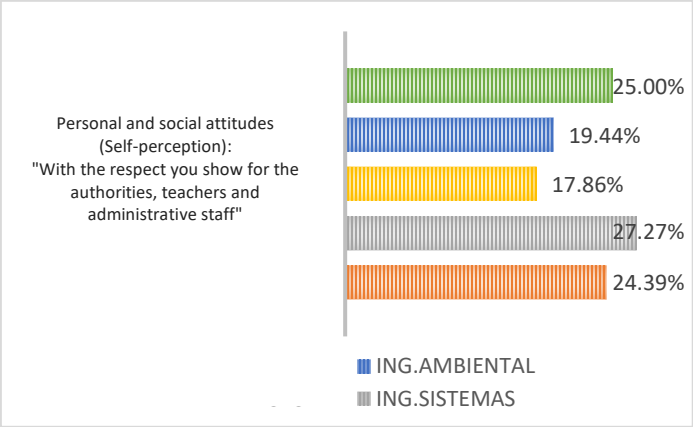


Figure 10: Indicators with the highest insatisfaction levels by specialty of the dimension personal and social attitudes (Self-perception)

As shown in Figure 10, the indicator that shows the highest percentage of dissatisfaction is "With the respect you show for the authorities, teachers and administrative staff" equivalent to 27.27% in the specialty of Systems Engineering in the Personal Attitudes and social (Self-perception).

4. Statistical Analysis

Initially for this analysis we will classify the dimensions into two groups dimensions by university services and self-perception dimensions of the graduate. This classification is shown in the following table.

Tabla 1: Clasification of dimensions

UNIVERSITY SERVICE DIMENSIONS	GRADUATE SELF-PERCEPTION DIMENSIONS
TEACHING PLAN	PROFESSIONAL COMPETENCES
UNIVERSITY ENVIRONMENTSD	PERSONAL AND SOCIAL ATTITUDES (Self-perception)
LIBRARY SERVICES	
LABORATORY SERVICES	
COMPUTER CENTER	
ADMINISTRATIVE SERVICE OF THE SCHOOL AND PROFESSIONAL SCHOOL	
UNIVERSITY SUPPORT SERVICES	

For this analysis we will see the correlation that exists between the indicators that present the highest percentage of dissatisfaction, for this we will be guided by the results obtained in the previous point; being our indicators to analyze: implementation of laboratories, records and enrollment, research skills and respect that students show towards the authorities, teachers and administrative staff; corresponding to the dimensions: laboratory services, University support service, professional skills and personal and social attitudes (self-perception), respectively.

To analyze the correlation you have the following interpretations:

There is no correlation = 0

Moderate positive correlation Strong +0.5

Perfect positive correlation +1.0

The results obtained in relation to the correlation of the indicators implementation of laboratories and research skills are shown below.

Table 2: Pearson correlation of the implementation of laboratories and research competences

		Laboratory Implementation	Competitions investigative
Laboratory Implementation	Pearson correlation	1	,651
	N	160	160
Competitions_ investigative	Pearson correlation	,651	1
	N	160	160

The above table shows a moderate positive correlation of 0.651, in relation to the dissatisfaction of the graduates towards having investigative competencies, because there was no adequate implementation for the professional development of the student of the National Technological University of Lima Sur.

Table 3: Pearson correlation the registrations and registration and personal and social attitudes

		Registrations and registration matriculas	Personal and Social Attitudes
Registrations_ and_ registration matriculas	Pearson correlation	1	,502
	N	160	160
Personal_and_ Social attitudes	Pearson correlation	,502	1
	N	160	160

The results obtained in relation to the correlation of the registration and enrollment indicators are shown below, and the respect shown by the students towards the authorities, teachers and administrative staff.

The above table shows a moderate positive correlation of 0.502, in relation to the dissatisfaction of the graduates towards the respect they show towards the authorities, teachers and administrative staff, this because the graduates during their years of study have always shown discomfort of the registration system, whose processes were carried out manually and the lack of technological tools was evident.

5. Discussion

In relation to the results of this research, the following discussions are held below:

- Regarding the indicators "Implementation of laboratories and the modernity of equipment and instruments" which show a level of dissatisfaction of 83.33% and 81.82%, respectively, it can be said that these agree with the study conducted by (Salinas & Martínez , 2017) which concludes that at the UPM, regarding the workshops where the students' practices are carried out; a highly dissatisfied degree of 30% is shown; This is reflected, because he did not have the necessary implements for the professional development of the student of the National Technological University of South Lima [10].

- Regarding the "Registries and enrollment" indicator which shows a level of dissatisfaction of 90.91%, it can be said that these agree with the study carried out by (Coveñas, 2017), in his thesis entitled Study of student satisfaction with services education provided by PROEDUNP - Sullana 2016, concludes that: Regarding administrative services; a dissatisfaction degree of 58.75% is shown; This is reflected, due to the delay in the enrollment process, since there was no enrollment system, which controls and optimizes the service of the students of the National Technological University of South Lima [11].

- Regarding the indicator "Availability of hygienic services" which shows a level of dissatisfaction of 81.82%, it can be said that they agree with the study conducted by (Coveñas, 2017), in his thesis entitled Study of student satisfaction with students educational services provided by PROEDUNP - Sullana 2016, concludes that: Regarding sanitary facilities; a dissatisfaction degree of 44% is shown; This is reflected, because there were cases of lack of cleanliness, as well as the availability of permanent drinking water in the hygienic services of the National Technological University of South Lima. [11]

6. Conclusions

- It is concluded that it was possible to determine the levels that present the greatest satisfaction of the graduate in relation to the quality of the educational services of their professional training at the National Technological University of South Lima, which are to assume a self-education (self-learning and continuing education) and I treat students in and out of the classroom; The two indicators present 85.71% satisfaction and belong to the dimensions of Professional Competencies and Teaching Plan respectively.

- It is concluded that it was possible to determine the highest levels that present dissatisfaction of the graduate in relation to the quality of the educational services of their professional training at the National Technological University of South Lima, which are Registries and enrollments with 90.91%, Implementation of the laboratories with 83.33% and availability of hygienic services with 81.82%; These indicators belong to the dimensions of University Support Services, Laboratory Services and University Environments respectively.

- Finally it is concluded that it was possible to determine the highest levels that present dissatisfaction of the graduate according to professional careers in relation to the quality of educational services in their professional training at the National Technological University of Lima, which are: in Electronic Engineering and Telecommunications, Services of Transportation with 87.80% and Availability of hygienic services with 63.41%, in Systems, Registries and Enrollment Engineering with 90.91%, Modernity of equipment and instruments with 81.82% and Implementation of laboratories with 81.82%, in Business Administration, Availability of hygienic services with 67.86% and Availability of books of their specialty with 57.14%, in Environmental Engineering, Implementation of laboratories with 83.33%, Presence of administrative staff in working hours with 66.67% and Availability of hygienic services with 75%, and in Mechanical and Electrical Engineering ica, Implementation of laboratories with 68.18%, The efficiency of the work of administrative staff with 65.91% and Internet service with 61.36%.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors wish to recognize and thank the Universidad Nacional Tecnologica de Lima Sur for their support of this investigation.

References

[1] Gonzáles, R., "Analysis of the satisfaction of the university experience of the graduates", Economic Paradigm, 59-84, 2016.

[2] Salinas, A., & Martínez, P., "Main satisfaction factors among university students", Revista Internacional de Ciencias Sociales y Humanidades, 163-192, 2017.

[3] Vázquez, M.; Hernández, M.; Mantecón, M. & Chávez, D., "Satisfaction of Cuban graduates of the Medicine degree with the training received", Curso 2015-2016, Cienfuegos. Universidad de Ciencias Médicas, Cienfuegos, Cienfuegos, Cuba, CP: 55100 Revista Medisur, vol.14: 512-513, 2017.

[4] Athiyaman, A., "Linking student satisfaction and service quality perceptions: thecase of university education", European Journal of Marketing, vol.31, núm.7: 528-540, 1997.

[5] Pérez Gil, J.A.; Lozano Lozano, J.A.; Gómez de Terreros Guardiola, M. Y Aguilera Jiménez, A., "Design of an instrument for the evaluation of the satisfaction of the training received from the different subjects corresponding to the Curriculum of the degree in psychology of the University of Seville", Revista de Enseñanza Universitaria, vol.36: 45-61, 2010.

[6] LapeñaLapeña, A. Y González Sánchez, M.C., "The continuous training of workers: manual del formador / [autores, Equipo IFES,]". Madrid Instituto de Formación y Estudios Sociales, 1996.

[7] González, M. &Sanchís, C., "Satisfaction of graduates with the training received in the Master of Advanced Studies in Business Administration", Journal of Education in Accounting, Finance and Business Administration.pp. 33-48, 2014.
[8] Paredes, M. &Cristaldo, M., "Satisfaction level of students and graduates of the National University of Asunción", Paraguayan Education Magazine, Volumen 1, Número 4, 2014, pp. 57-69, 2013.
[9] Landa, M. & Ramírez, M., "Design of a student satisfaction questionnaire for a profesional level course under the inverted learning model", Revista Páginas de Educación. Vol. 11, Núm. 2, 2018.
[10] Salinas, A. M., & Martínez, P. "Student satisfaction and university quality: an explanatory analysis in the academic and multidisciplinary agronomy and science unit of the Universidad Autónoma de Tamaulipas", México. Revista de Enseñanza Universitaria, 39-65, 2008.
[11] Coveñas, J., "Study of student satisfaction with the educational services provided by PROEDUNP-Sullana 2016", UNIVERSIDAD NACIONAL DE PIURA, Piura, 2017.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 158-162 (2019)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Android Application to Detect Cat Disease Using an Expert System

Alvina Aulia[1], Yogi Udjaja[1,2,*], Irma Kartika Wairooy[1], Adithia Panji Hutama[1], Dinar Khansa Shabira[1], Sabilal Muhtadin[1]

[1]*Computer Science Department, School of computer Science, Bina Nusantara University, Jakarta, Indonesia 11480*

[2]*Ekspanpixel, Jakarta, indonesia 11480*

ARTICLE INFO

Article history:
Received: 26 April, 2019
Accepted: 19 July, 2019
Online: 24 September, 2019

Keywords:
Cat
Diagnosis of Cat Disease
Android Application

ABSTRACT

The purpose of this research is to develop an android application named "TANYA MEOW" which can be used by cat owners to detect the diseases that their cats suffer based on visible symptoms. The cat owners need to answer questionnaire to detect what disease is suffering by their cat. The result of the disease is also accompanied by the first treatment that cat owners can do to treat their cats. In addition, this application also has a feature that provides information, tips and tricks in cat care. The research methods used to analyze the development of this application are literature study through books, journals and e-books. Furthermore, observation, distribution of questionnaires for data collection and interviews to ensure data accuracy. We also make comparisons with applications that have similar concepts and features similar to our application. The research method used for the development of this application is waterfall that has five stages, requirements definition, system and software design, implementation and unit testing, integration and system testing and operation and maintenance. The results obtained from this writing show that the owner of the cat can feel the ease in detecting and knowing the disease from their cats who are sick, also getting knowledge on how to take care of cats in a good way.

1. Introduction

In the era of rapid development of smartphones, where pet owners are now so enthusiastic to see the behavior of their pets and capture this moment through the camera on the smartphone. Animal owners also often use smartphones to access various articles on the internet to recognize the nature, behavior, and symptoms of the disease [1]. However, some articles are not actual and true for the health of their pets. Animal owners need a good source and useful media to see the condition of their pets. Thus, we developed an app that focused on cat that could allow users to find ways to prevent and perform first action on their beloved cat when strange symptoms seen in cats or cats are already infected by an illness as well as providing tips and tricks in the form of actual information in the care of cats [2]. Therefore, this application can be one of an application or a media to help cat owners in doing the right action based on the illness suffered in the form of useful and actual information for their cat [3].

Therefore, the formulation of our problem in making this research is how cat owners can know the diseases suffered by cats based on visible symptoms, how cat owners can find information about handling the disease, first aid on what to do when the cat is sick, how cat owners can find information about the prevention of diseases that must be done when the cat is sick and how cat owners can find out information about tips and tricks in cat care using android based smartphones [4]. From the problems in, it can be concluded that our goal in making this application is to create an application that can help cat owners to know the type of cat disease based on visible symptoms, provide information that can help cat owners to know easily about the handling and first aid against cat, providing information that can help cat owners find out easily about cat's disease prevention and provide information on android smartphones that can help cat owners find out easily about tips and tricks in cat care [5]. The benefits of the for cat owners to rapidly diagnose cat diseases, accurately and interactively, help cat owners to take early action against diseases suffered by cats, provide detailed information on cat disease prevention and increase knowledge to owners cat about tips and tricks in cat care.

2. Why Cat?

Cats are animals that have many advantages. His fur is soft, his body shape is adorable, his spoiled nature always makes him love to stroke and play with him. In addition, the behavior of cats

*Yogi Udjaja, Jakarta, Indonesia 11480, +62 898 268 3399, udjaja.yogi@gmail.com, yogi.udjaja@binus.ac.id

that are very funny while joking makes their lovers entertained. Initially, cats come from the wild, then slowly undergo the process of dosmetication. Now, cats become pets that are very close to humans [6].

According to ref. [7], the following are examples of diseases that can be suffered by cats:

- Cat Diarrhea

This disease is a disease of increased fluid in the stool and increased frequency of defecation. Mild diarrhea is usually caused by changes in food, too much eating, and stress disorders. Conversely, severe diarrhea is caused by an increase in peristaltic movements of the intestine so that the food material comes out faster, food does not have much time to digest properly.

Clinical symptoms include: runny diarrhea, very runny stools. Cat feces smell bad. The cat vomits and refuses to drink. To prevent diarrheal disease, clean the place to drink and place food with disinfectants. For treatment you can use worm medicine and can be injected with vitamin B12 to increase appetite.

- Cat Lice

Lice in cat are caused by Ctenocephalides felis. Adult fleas can live up to around 115 days in a cat's body, and will die soon if they are not in the cat's body for only 1-2 days. The survival of adult lice depends on environmental conditions, including humidity, cat's breath, and cat's body vibration.

The initial symptoms of cat lice are just normal itching. All fleas will irritate your cat's skin. Sometimes because of the large number of fleas and itchiness, cats scratch their bodies, especially in the nape of the neck, ears and base of the tail. Prevention and treatment efforts can be with flea drugs and with lice exterminator shampoo.

- Cat Mite

This disease is generally caused by Otodectes cynotis. The life cycle from eggs to adults takes around 21 days. Mites often attack young cat ears. They eat secretions in the ear canal or flaking skin. The initial symptoms of cat mite are shaking their heads and scratching their ears, sometimes cats rub their ears, walls, or soil. For symptoms that are severe the cat bends and loses balance and sometimes rolls over. In the ear hole there is dry or damp blackish brown dirt and thickening of the skin.

Symptoms can continue with the release of a runny or thick liquid from the foul-smelling cat's ears and yellow to greenish red. Prevention efforts that can be done is by washing or changing the cat's bed, every 3-4 weeks the cat's ear is given drops, mite exterminator, cleaning the hole.

- Cat Otitis

This disease is an infection of the ear canal, which is a connecting area between the outside of the ear and ear drum. Symptoms of itching in the ear. The cat tries to shake its head, scratch, and sometimes rubs the ear against the cage wall until a wound appears, the ear hole swells, and the cat feels pain when the part is held. The skin of the inner ear is reddish, the skin will thicken and scaly, pus and bumps sometimes appear.

Prevention that can be done is by eradicating ticks and fleas that attack the cat's cage. Clean cat ears once a week. Treatment can be adjusted to the cause. For bacteria given antibiotics, for fungi given antifungals, for allergies can be given antihistamine drugs.

- Cat Constipation

The first cat constipation is found in food quality. Cats are very short of crude fiber and drink less, including eating too much meat. Another cause is the age of the cat, where the intestinal muscles have started to weaken so that the activity decreases.

The symptoms are difficulty defecating Anus swells, reddish dirt and sometimes accompanied by blood and during bowel movements, cats always try to push, but dirt fails to come out. Prevention that can be done is to provide a balanced diet that is sufficiently like rough fiber. Always provide clean and fresh water so the cat is not short of fluids. The first treatment for mild constipation is to give laxative medication. Use laxatives to remove hard feces quickly.

3. Problem Statement and Preliminaries

The first research method used is data collection methods, including literature study conducted to find sources and references related to the development of Android-based mobile applications. Second, observations made by observation to animal clinics. Observation is done by recording the implementation of activities in the field, especially at the time of the interview. Third is the distribution of questionnaires conducted as a method of data collection. Fourth is interview conducted to the veterinarian to ensure the truth of data about the disease as well as symptoms in the cat and asking information that is needed to be implemented in the application so that the data obtained are actual. Then an application is formed using an expert system [8]. The last is the comparison of applications performed by taking reference applications that have the same characteristics and within the same category that will be an important source for the progress of this research.

4. Analyze User Needs

In the process of collecting data for the needs of users, we used questionnaire method. The questionnaire are distributed to 124 respondents on September 11, 2017 to September 14, 2017. Here are the results of the questionnaire:

1. 100% of respondents have cats.
2. Of the total 100%, 39.5% have domestic cats, 33.9% have Persian cats, 6.5% have cats, 2.4% have Maine Coon cats and 17.7% have cats other races.
3. 78.2% of respondents keep cats inside their house and 21.8% keep outside their house.
4. 73.4% of respondents answered that their cats rarely get sick, 21% are quite frequent and 5.6% are often sick.
5. 41.1% of respondents were browsing about diseases and handling it by their own when their cats were sick, 35.5%

were taken to the doctor, 15.3% asked the closest relation that had a cat and 3.2% answered to be silenced until healed by itself.

6. Of 100% of respondents as much as 83.9% responded confused when their cat changed behavior and as much as 16.1% answered "No".
7. 38.5% of respondents know the disease in cats and symptoms and as much as 61.5% do not know the disease and the symptoms in cats.
8. With a total percentage of 100%, 96% are interested in using application that can detect disease, provide tips and tricks and first aid information on cats.

Table 1: Similar Application Comparisons

Comparison	Cat care	Pet care
Platform type	Android	Android
Internet	Requires internet	No need for internet
Notification	Do not have tutorial	Do not have tutorial
Article	Does not have article features	Has article features
Background Music	Do not have background music	Do not have background music
Disease Diagnosis	Does not have a diagnosis of disease features	Does not have a diagnosis of disease features
Handling Disease	Does not have a disease handling feature	Does not have a disease handling feature
User Interface	Easy to understand interface and interactive menu	Simple and easy to understand interface
Advantages	Contains news about cats	Contains articles, tips and tricks about pets, including cats

5. Similar Application Comparison Analysis

Comparison of similar applications is adopted from ref. [9], this model is a comparison of features to the advantages possessed by each application analyzed, so that the application made can be developed in accordance with the results of comparison of similar applications.

Based on comparison of similar applications, the target is the Android platform, to make it easier for users of the application to not need the internet, to have tutorials, to feature articles, and to make music background interesting, then to diagnose diseases, to treat diseases and to make the user interface easy to understand, interesting and interactive. So overall the advantages possessed compared to other applications are having a feature of disease diagnosis, disease management and containing articles about tips and tricks in maintaining cats.

6. Architecture Software

Architectural models that are made almost similar to architectural ref. [10], what is different is the content in it, where there is an expert system to detect cat disease. Then, for the user experience, use the model created by ref. [11, 12].

7. Software Functions

1. Cat's Registration
   This feature is used for cat registration by filling in data on registration form such as name, gender, month of birth and race. Users can register one cat or more.
2. Tanya Meow (Diagnosis of Disease)
   This feature is used to diagnose the disease in cats based on visible symptoms. First, the user will choose what parts of the body that cause symptoms of the disease. Then, users are given some questions based on these symptoms and given the diagnosis of a cat-like disease and the first aid on what to do to treat their cat.
3. Info Meow
   This feature contains information, articles and tips and tricks how to care for a good and true cat. The menu is divided per-category for easy user in navigating.
4. P3K
   This feature contains treatment of the diseases that can be suffered by cats. The menu is divided per-category for easy user in navigating.
5. Cat Disease's History
   This feature is used to view the list of selected cat catastrophes that users have previously done in the menu "Tanya Meow".

8. Main Results

Below is the display of features or app results of *Tanya Meow:*

- Fig.1 The main menu contains the Registrasi, Pilih Kucing, Tanya Meow, Info Meow, P3K, Riwayat, and Pengaturan menu. This page provides information to users about what features can be used on the main menu page.
- Fig.2 This page contains forms to fill with the names of cats, gender, month of birth, race of cats, and also choose the cat icon. The cat registration page helps the user create a cat profile. On the Pilih Kucing page, the user can see the registered profile and choose the desired cat profile. This page allows the user to choose the cat that will be used.
- Fig.3 Choosing a cat profile or deleting a pre-registered profile can be done on the Konfirmasi Pilih Kucing page. This page helps the user to make choices based on the user's wishes.
- Fig. 4 Tanya Meow page is a diagnostic page that helps the user to know the cat's disease.

- Fig 5. In the Pertanyaan Tanya Meow page, user can choose the answer option based on the given question by choosing "Yes" or "No" button until Pertanyaan Tanya Meow page result appears.
- Fig. 6 Hasil Tanya Meow page is the result page of the answer to the question the user chose. This page provides information about the results of diseases suffered by cats based on the answers to questions that users choose.
- Fig. 7 Info Meow page is a page that displays articles that contain information on how to take care of cats. This page allows users to get articles quickly because they are organized by certain categories.
- Fig. 8 P3K is a page that displays articles that contain information about the first handling and prevention of cat diseases. This page provides users with information about what actions a user can take when their cat is sick.
- Fig. 9 Riwayat page will show a list of cat diseases that have been diagnosed. This page allows users to see the diagnosis of the disease that has been done.

9. Evaluation

After produced “Tanya Meow” application, The evaluation was done by distributing questionnaires to 30 respondents (users) based on the measurable 5 human factor theory and the conclusion that 91% users are satisfied with the features of the “Tanya Meow” application and assisting them in knowing cat's disease, adding insight on cat care and prevention and first aid to do when the cat is sick.

10. Evaluation

Based on analysis and evaluation conducted, the conclusion is as follows: by using “Tanya Meow” application, users can easily know the illness of their ailing cat based on the symptoms seen, get proper handling and first aid on their sick cat, can help users to get information about disease prevention should be done by the owner of the cat. For future work, there are some suggestions that need to be added to the “Tanya Meow” application as the following: Make data statistics about diseases that are often infected by cats based on users of Tanya Meow application that is connected online, and provide question and answer features with veterinarians and provide information on experienced veterinarians.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] C. Hargrave. “Behavioural first aid for elderly cats and dog”. Veterinary Nursing Journal, 26(11), 410-412, 2011.
[2] T. J. Howell, K. Mornement, P. C. Bennett. “Pet cat management practices among a representative sample of owners in Victoria, Australia”. Journal of Veterinary Behavior: Clinical Applications and Research, 11, 42-49, 2016.
[3] T. V. Magazzu. “Pet Talk - First Aid”. Courier Journal, 2014.
[4] S. J. Russell, P. Norvig. “Artificial intelligence: a modern approach”. Malaysia; Pearson Education Limited, 2016.
[5] A. V. Saurkar, H. N. Watane. “An Expert Syatem For Diseases Diagnosis in Pet”. Advances in Medical Informatics, India, 2012.
[6] M. A. Suwed, R. M. Napitupulu. “Panduan Lengkap Kucing”. Penebar Swadaya Grup, 2011.
[7] H. S. Wheirendrata. ”Buku Pintar Kesehatan Kucing Ras”. Yogyakarta: Lily Publisher, 2016.
[8] Y. Udjaja. “ANDROID APPLICATION FOR DETECTION OF SKIN CANCER USING EXPERT SYSTEM”. Social Economics and Ecology International Journal (SEEIJ), 2(1), 2018.
[9] Y. Udjaja, V. S. Guizot, N. Chandra. “Gamification for Elementary Mathematics Learning in Indonesia”. International Journal of Electrical and Computer Engineering, 8(5), 3859, 2018.
[10] Sasmoko, Y. Indrianti, S. A. Widhoyoko, Y. Udjaja, A. Tanurwijaya. “Teacher Engagement Interventions through ITEI Apps”. In 2018 6th International Conference on Cyber and IT Service Management (CITSM) (pp. 1-4). IEEE, 2018.
[11] D. P. Kristiadi, Y. Udjaja, B. Supangat, R. Y. Prameswara, H. L. H. S. Warnars, Y. Heryadi, W. Kusakunniran. “The effect of UI, UX and GX on video games”. In 2017 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom) (pp. 158-163). IEEE, 2017.
[12] Y. Udjaja. “Ekspanpixel Bladsy Stranica: Performance Efficiency Improvement of Making Front-End Website Using Computer Aided Software Engineering Tool”. Procedia Computer Science, 135, 292-301, 2018.

Figure 1: Main Menu Page

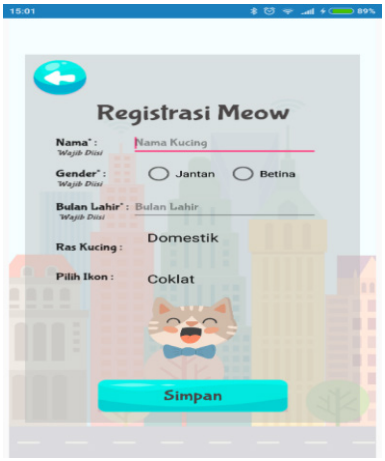


Figure 2: Registration Page

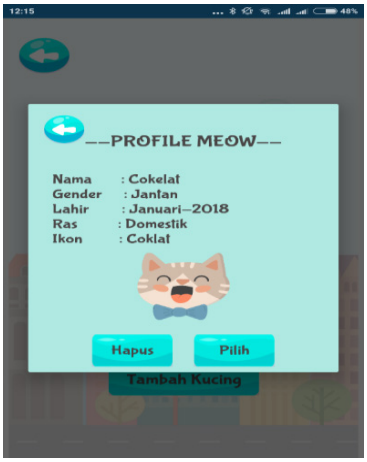


Figure 3: Cat's Profile Page

Figure 4: Tanya Meow Page

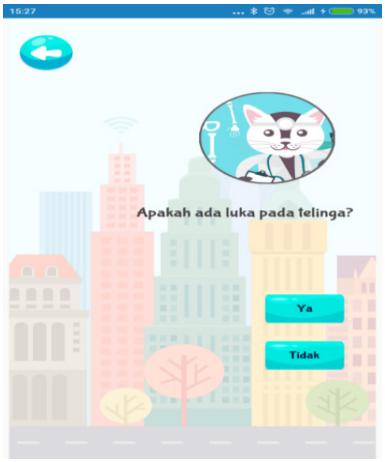


Figure 5: Question and Answer Page

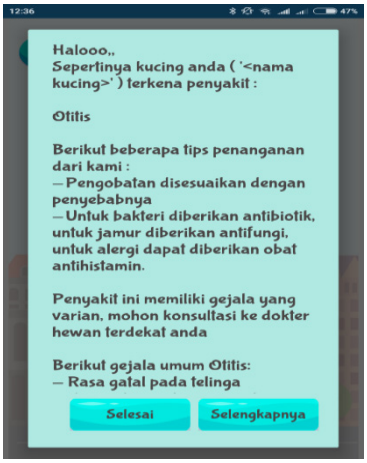


Figure 6: Result Page

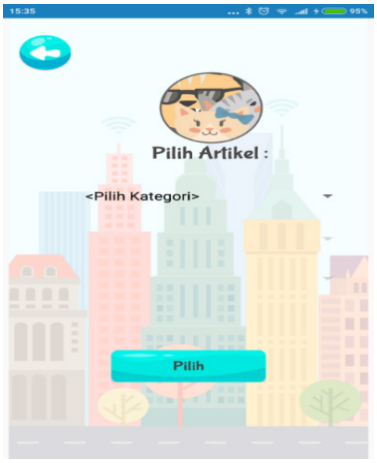


Figure 7: Article Page

Figure 8: P3K Page

Figure 9: History Page

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 5, 163-173 (2019)
www.astesj.com
Special Issue on Multidisciplinary Sciences and Engineering

ASTES Journal
ISSN: 2415-6698

Chatting with Plants (Orchids) in Automated Smart Farming using IoT, Fuzzy Logic and Chatbot

Samruan Wiangsamut, Phatthanaphong Chomphuwiset, Suchart Khummanee*

Department of Computer Science, Faculty of Informatics, Mahasarakham University, 44150, Thailand

ARTICLE INFO

Article history:
Received: 16 August, 2019
Accepted: 25 September, 2019
Online: 10 October, 2019

Keywords:
Automated smart farming
Internet of Things (IoT)
Fuzzy logic
Chatbot

ABSTRACT

Plants are living organisms that can hear and recognize the environment around them but cannot communicate to inform their needs. Thus, in the past, humans thought that it was impossible to communicate with plants. However, in this modern era, humans can be able to communicate with these plants. In this paper, we propose a model that can interact (Chat) with plants cultivated in the automated farm system based on Internet of Things (IoT) and Fuzzy Logic. According to the communication of plants and humans, we apply a chatbot algorithm for sending/receiving messages between users and automated smart farming. The messages are processed by the natural language processing (NLP) to parse and interpret the meaning of the conversation. The experimented plant in this paper is orchid, namely Dendrobium Sonia (Bomjo). The result from the evaluation shows that the average accuracy (Harmonic mean) of chatting between the user and the orchid is equal to 0.71, the precision and recall are 0.75 and 0.6 respectively.

1 Introduction

What happens if we can talk to trees or plants? The answer is that we can know what they want, which will help us to respond to their needs accurately and precisely. For example, if the temperature rises, the moisture in the soil is decreased; as a result, plants need more water. If they are not watered at the right time, they will gradually wilt and eventually die. Besides, there is another exciting thing about plants, that is, the sound waves of different frequencies can affect the growth of different trees [1, 2]. Fahmi Huda et al. [3], for instance, discovered that acoustic waves could help stimulate the absorption of plant nutrients more efficiently, resulting in height, leaf area, stomatal opening, chlorophyll content, absorption of nitrogen, etc. of mustard Pakchoy are also higher.

The basic principle of chatting between plants and humans is applied and derived from the communication between ordinary and voiceless people. The deaf-mute can perceive various environments by looking, hearing, taste and touching, except for pronunciation, which is similar to the nature of plants. In fact, plants can recognize the various dangers that are approaching them, but they cannot response to aggressors. For example, the Mimosa pudica will shut their leaves if they are vibrated, or Morning glory flows will bloom or fold as the light intensity, etc. Therefore, we have adapted the principles mentioned above to apply to our chatting technique. Several researchers contributed methods for controlling and ordering various devices using the chat method via chatbots. Cyril Joe Baby et al. [4], for example, proposed a way to control various electrical devices in homes using a chatbot, working with IoT, that allow users to control devices within their house in a manual only. The system has not yet evaluated the errors regarding conversations via a chatbot and is not yet automated. Similar to Cyril's method, Muslih [5] introduced techniques for controlling electrical appliances in the workplace with IoT via Telegram Chatbot program in a smartphone. They claimed that the system can handle appliances correctly and also reduce the energy used, but they did not show the pros and cons that occurred from the communication between the IoT controller and the chatbot program. Besides, Angeline et al. [6] apply a chatbot implemented by Python to chat with customers when they need to buy goods in supermarkets, but it is only an extension, not an essential part of the system. Chatbots are applied to several ways [7], for example, Kaed et al. [8] developed an integrated system which collects data obtained from multiple IoT platforms running in the background, with a chatbot acting as a communication against the users in the foreground. Such a system is an intermediary that allows different IoT platforms to communicate with each other via a chatbot.

*Suchart Khummanee, Department of Computer Science, Faculty of Informatics, Mahasarakham University, Thailand. Email: suchart.k@msu.ac.th

In this paper, we propose techniques to chat with plants (Orchids) cultivated in the automated smart farming. The proposed techniques are based on IoT and Fuzzy logic using natural language processing (NLP) via the Python chatbot program. This paper is an extension of work initially presented in 2018 International Conference on Information Technology (InCIT) [9].

The rest of the paper is organized as follows: Section 2 depicts the background and related work. We articulate our methodology in section 3. The performance evaluation shows in the section 4. The system development is described in section 5. Section 6 shows our experimental results, and the last is the summary of this paper.

2 Background and Related Work

2.1 Stimulation and Recognition of Plants

Naturally, plants can recognize the stimulus from both inside and outside their bodies, which is a mechanism that occurs from the functioning of hormones or the various mechanisms of cells to make plants grow [10]. The stimuli influence the response of plants, which can be divided into two groups: external and internal stimuli. Plant's responses are different depending on these factors, such as the duration of stimulation of stimuli, the age and cell types that the sensory unit (Receptor), and the amount or concentrations of stimuli, etc. When plants are stimulated by stimuli, they can respond in several ways, such as movement, chemical release and tissue changes. For example, the leaves of Mimosa will be folded if there is a vibration near their leaves, as shown in Figure 1 (a). In this research, we are only interested in orchids because they are economic crops and have a high export value. Orchids (Dendrobium shown in Figure 1 (b)) can grow well if the natural stimulation factors: sunlight is 60 percent, the temperature is between 25 - 35 degrees Celsius, the relative humidity is between 60 - 70 percent and soil moisture is in the range of 60 - 70 percent [9].

(a) Mimosa (b) Orchid (Dendrobium)

Figure 1: Specimens of Mimosa and Orchid

2.2 Smart Farming with IoT

Smart farming is modern agriculture, which works with new technologies (IoT), robots, machines, etc. It has high accuracy for resolving problems of farmers by being friendly to the environment, being safe for consumers and using resources efficiently. For instance, Anukit and et al. [11] developed an alternative electric power supply from solar energy for a smart agricultural greenhouse, called the smart photovoltaic system, which is applied to the Lingzhi mushroom farm for saving energy costs. Besides, they [9, 12] used IoT to control the environment for the growth of orchids inside a smart farm due to unstable weather conditions, which allows farmers to plan production precisely. Hydroponics is another popular method today because farmers can get vegetables that are safe and easy to grow. Boonnam et al. [13] proposed a method of growing vegetables on a hydroponic farm controlled by IoT via mobile phones. The overall evaluation result from users' confidence in their system is higher than 70%. Next, farmers can make accurate decision making while undesirable events occur inside their farms based on real-time data using Agri-IoT [14] running on the semantic web technologies. IoT is not only restricted to use for crops, but it is also used for animal farms. For example, Memo et al. [15] adapted IoT to monitor various environments inside an animal farm, such the animal's waste, fire, water, animal's behaviours and others via a smartphone. The interesting researches related to IoT and the smart farming: [16, 17, 18].

2.3 Automatic Smart Farming with Fuzzy Logic

The automatic smart farm is a farm which can control various environments to suit each plant within the farm by itself without human intervention. For example, orchids grow well at a relative humidity (humid) ranging from 60 to 70% If humid around the farm is now 59%, therefore, the automatic smart farm must control the humid back to 60 to 70% by fogging automatically. To make any farm to be automated, there are a variety of methods, for a prevalent method, that is Fuzzy Logic (FL) [19]. Because of FL can explain ambiguous information so that humans can understand, especially the various environment information of plants. For example, three farmers choose unequal orchid growing temperatures, such as 25 - 30°C, 26 - 32°C and 25 - 32°C respectively. Therefore, these ambiguous data will be calculated using Fuzzy Logic. As a result, the appropriate temperature data range for orchid cultivation is 25 - 30°C. The output evaluated by the FL from multiple input levels is finite. There are four steps, as shown in Figure 2.

1. **Fuzzification** translates inputs, called crisp inputs into Fuzzy sets, and splits each crisp input into multiple levels, i.e., five levels = NH (Negative High), NL (Negative Low), Z (Zero), PL (Positive Low) and PH (Positive High) or three levels = L (Low), M (Medium) and H (High),

2. **Knowledgebase** creates rules that correspond to the problems being considered in the IF-THEN format, for example, IF $\text{humid}_{present}$=L AND humid_{Target}=M THEN Fogging,

3. **Inference engine** simulates human reasoning by using Fuzzy inference and knowledge base in conjunction with inputs,

4. **Defuzzification** transforms the Fuzzy set acquired from inference engine into a crisp output.

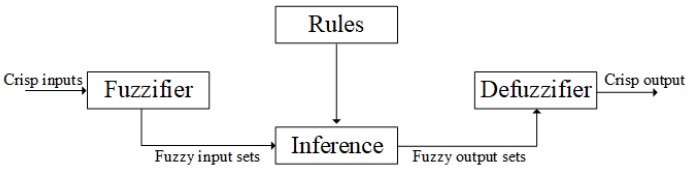


Figure 2: Fuzzy Logic workflow

2.4 Chatbots

A chatbot or virtual assistant is an artificial intelligence program [4, 7] which can simulate a conversation (Chat) with a user in natural language through messaging applications, websites or smartphones. Chatbot's structure consists of two parts: user request analysis and returning the response, as shown in Figure 3. Generally, chatbots work on three classification methods:

1. **Pattern Matching**: chatbots use a pattern matching method to group messages and create relevant responses from users. Artificial intelligence Markup language (AIML) is a standard structured model to create an interface which can be easily understood by humans response to users with the appropriate format, as shown in Figure 4.

2. **Natural Language Understanding (NLU)**: NLU is a subset of Natural Language Processing (NLP) which examines the group of texts or speeches (Sentences) and interprets the meaning specified. For example, a query for asking temperature inside Orchid smart farming on the 20th of July. NLU algorithm breaks down the query something like this: Orchid [intent] / need: temperature [intent] / inside farm [location] / July 20th [date].

3. **Natural Language Processing (NLP)**: NLP uses algorithms to analyze text so that the machine understands how humans speak. It is commonly used for text mining, machine translation, and automated question answering. NLP uses the machine learning algorithms which has the following steps: Tokenization, Sentiment Analysis, Normalization, Named Entity Recognition and Dependency Parsing, respectively [20].

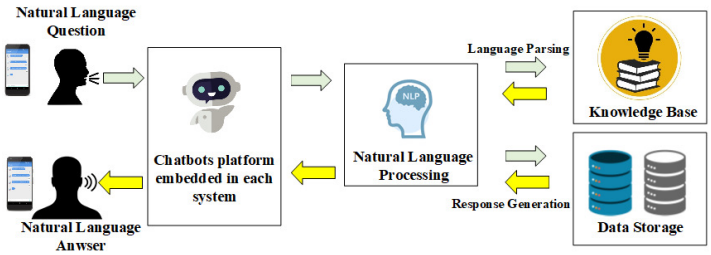


Figure 3: Chatbot's workflow

```
<aiml version="1.0" encoding ="UTF.8"?>
<category>
<pattern>HI HOW ARE YOU</pattern>
<template>
<srai>HI<srai><srai>HOW</srai><srai>ARE YOU</srai>
</template>
</category>
<category>
</aiml>
```

Figure 4: Artificial Intelligence Markup Language(AIML)

3 Proposed Methodology

Figure 5 illustrates the research procedures in order to build a chatbot for automated smart farming of Orchids. There are eight steps (1 to 8) as:

3.1 Collect information about orchid cultivation from farmers

The first step (1), we collect as much information as possible about the environments which are used to plant orchids from farmers including temperature (T), humidity (H) soil moisture (S) and sunlight (L). Information collected from orchid farmers in this research has eight sources listed in Table 1.

Table 1: Information on orchid cultivation from Orchid farmers

Farmer	Environment variables			
	Temperature (T)	Humidity (H)	Soil moisture (S)	Sunlight (L)
$\mathbf{F}_1$	25-30°C	N/A	N/A	N/A
$\mathbf{F}_2$	25-35°C	60-80%	60-80%	60-70%
$\mathbf{F}_3$	25-35°C	60-80%	60-80%	50-60%
$\mathbf{F}_4$	N/A	60-80%	N/A	N/A
$\mathbf{F}_5$	25-35°C	N/A	N/A	N/A
$\mathbf{F}_6$	25-35°C	50-70%	50-70%	N/A
$\mathbf{F}_7$	N/A	N/A	N/A	50-60%
$\mathbf{F}_8$	25-30°C	N/A	N/A	50-70%

Where F_i as any farmer, and N/A is unavailable.

According to Table 1, the first farmer (F_1) recorded only temperature (T) ranging from 25 to 30 degrees Celsius. Other information, the farmer does not record (N/A). In contrast, the farmer F_2 and F_3 recorded all the information of the environment variables: T ranges from 25 - 35 °C, H and S are between 60 - 80%, L is in the range of 60 - 70% and 50 - 60% respectively. Each type of information is calculated to determine membership values by the intersection operator ($\cap$), except the N/A as follows.

$$\begin{aligned} T_m &= T_{F_1} \cap T_{F_2} \cap T_{F_3} \cap T_{F_5} \cap T_{F_6} \cap T_{F_8} \\ &= [25, 30] \cap [25, 35] \cap [25, 35] \cap [25, 35] \cap \\ &\quad [25, 35] \cap [25, 30] = [25, 30] \end{aligned} \tag{1}$$

T_m is an intersection result of T_{F_i}, where i $\in$ [1, 8].

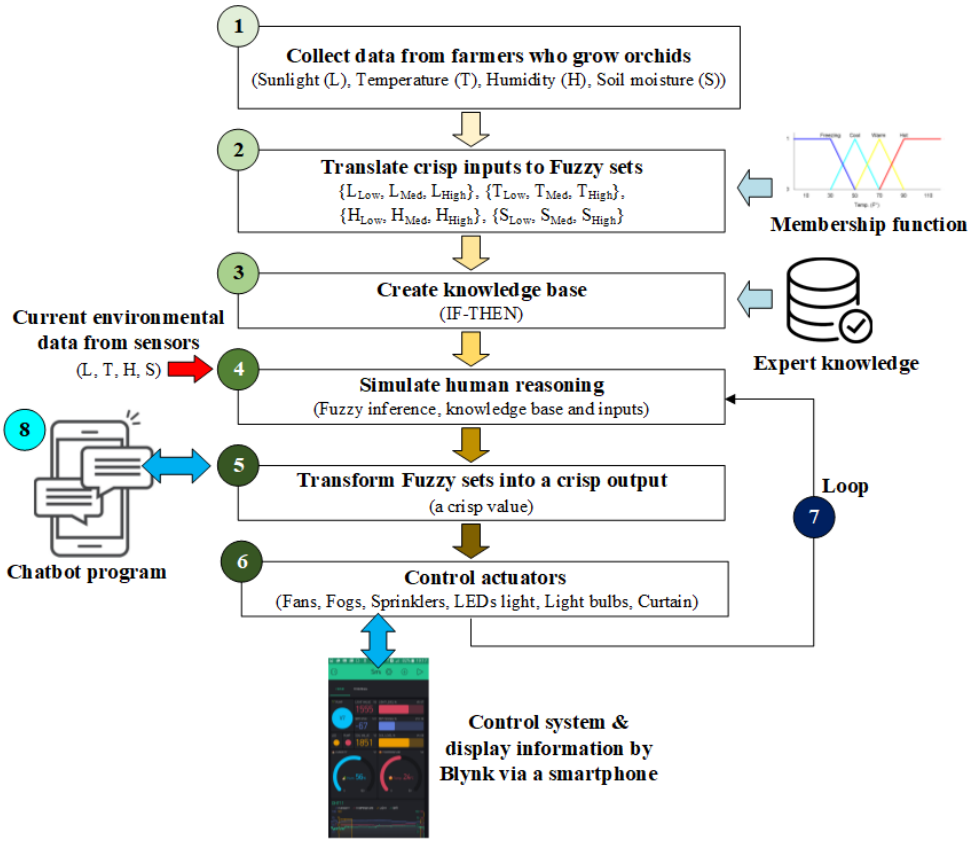


Figure 5: Block diagram for research methodology

$$H_m = H_{F_2} \cap H_{F_3} \cap H_{F_4} \cap H_{F_6} = [60, 80] \cap [60, 80] \cap [60, 80] \cap [50, 70] = [60, 70] \quad (2)$$

$$S_m = S_{F_2} \cap S_{F_3} \cap S_{F_6} = [60, 80] \cap [60, 80] \cap [50, 70] = [60, 70] \quad (3)$$

$$L_m = L_{F_2} \cap L_{F_3} \cap L_{F_7} \cap L_{F_8} = [60, 70] \cap [50, 60] \cap [50, 60] \cap [50, 70] = [60] \quad (4)$$

H_m, S_m and L_m are the intersection results of H_{F_i}, S_{F_i} and L_{F_i}, respectively.

3.2 Translate crisp inputs to Fuzzy sets

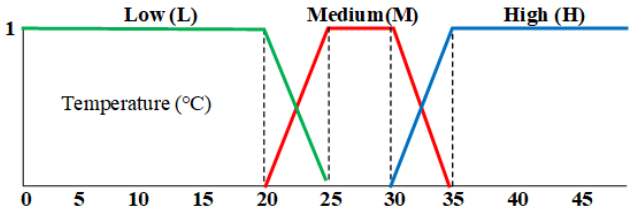
The second phase (2), crisp inputs are converted to Fuzzy sets, which are obtained from the first step. Theoretically, crisp inputs can be multiple levels, but in this research, only three levels are selected: T, H, S and L ∈ {LOW, MEDIUM, HIGH} because orchids are not very sensitive to climate changes. The membership functions are used to map crisp inputs to Fuzzy sets. They quantify linguistic term and demonstrate Fuzzy set graphically. A membership function for a Fuzzy set T on the universe of discussion X is defined as μT:X $\longrightarrow$ [0, 1]. Here, each element of X is mapped to a value between 0 and 1. Each membership function of T, H, S and L are illustrated in Figure 6 (a to d) respectively.

3.3 Create knowledge base rules

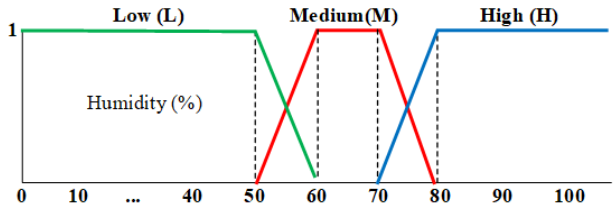
Next step, we map each membership function from the second step to a matrix, which is determined how we proceed if environment variables in smart farming occur in different situations. For example, from Figure 6 (a), if a current temperature ($T_{current}$) inside a smart farm is now equal to 24°C, but a target temperature (T_{target}) for planting orchids is in the middle (T ∈ [25, 30°C] = M); thus, the system needs to immediately reduce the current temperature to be the middle level (M) as shown in Table 2.

Another example, if the current relative humidity ($H_{current}$) is low about 58%, but the humidity suitable (H_{target}) for orchid planting is 60 - 80% (Middle). Therefore, the system will have to carry out fogging (Fog) to increase the humidity at an appropriate level immediately shown in Table 3. For other environment variables as S and L are mapped into a matrix, shown in Table 4 and 5 respectively.

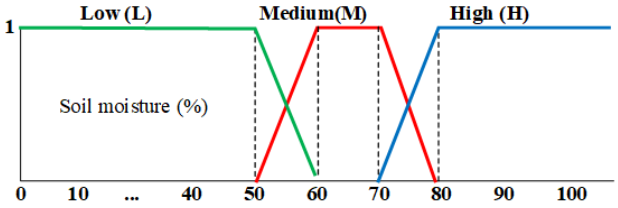
Next, we build a set of rules of T, H, S and L of each matrix in order to build the knowledge base in the form of IF-THEN structure like in Table 6. According to Table 6 rule No. 4, if the current humidity in the farm is high (H) and the orchids need moderate humidity (M), the system must turn on lamps to reduce the humidity inside the farm, and order the fans to suck moisture away from the farm simultaneously.



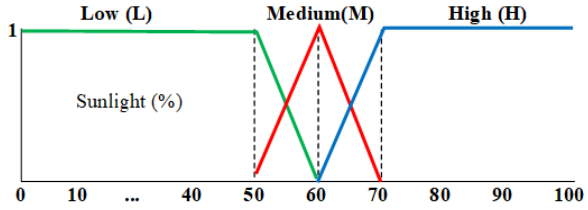
(a) Temperature (T)



(b) Humidity (H)



(c) Soil moisture (S)



(d) Sunlight (L)

Figure 6: Membership function of each crisp input

Table 2: The decision matrix of temperature

$\mathbf{T}_{target}$	$\mathbf{T}_{current}$		
	Low	**Medium**	**High**
Low	No change	Cool	Cool
Medium	Heat	No change	Cool
High	Heat	Heat	No change

T_{target} = target temperature for planting orchids, $T_{current}$ = current temperature inside a smart farming

Table 3: The decision matrix of humidity

$\mathbf{H}_{target}$	$\mathbf{H}_{current}$		
	Low	**Medium**	**High**
Low	No change	Heat/Fan	Heat/Fan
Medium	Fog	No change	Heat/Fan
High	Fog	Fog	No change

H_{target} = target humidity for planting orchids, $H_{current}$ = current humidity inside a smart farming

Table 4: The decision matrix of soil moisture

$\mathbf{S}_{target}$	$\mathbf{S}_{current}$		
	Low	**Medium**	**High**
Low	No change	No water	No water
Medium	Water	No change	No water
High	Water	Water	No change

S_{target} = target soil moisture for planting orchids, $S_{current}$ = current soil moisture in planting pots

Table 5: The decision matrix of sunlight

$\mathbf{L}_{target}$	$\mathbf{L}_{current}$		
	Low	**Medium**	**High**
Low	No change	Curtain	Curtain
Medium	Lighting	No change	Curtain
High	Lighting	Lighting	No change

L_{target} = target lighting for planting orchids, $L_{current}$ = current lighting inside a smart farming

3.4 Simulate human reasoning

Fuzzy set operations evaluate rules created in the previous step (Section 3.3). AND ($\wedge$) and OR ($\vee$) are used for describing a minimum (MIN) and a maximum value (MAX) respectively. Combine all results of the evaluation to form a final result, called a Fuzzy value. For example, the membership function of temperature shown in Figure 7 has three output Fuzzy sets: "Low", "Medium" and "High". Assuming that the temperature in the current orchid farm is equal to 23 °C, thus the membership value of the Fuzzy set "Low, Medium and High" are 0.35, 0.75 and 0, respectively. The Fuzzy set value is 0.35 $\vee$ 0.75 $\vee$ 0 = 0.75 (MAX). As a result, the Fuzzy value is Medium (M).

3.5 Transform the Fuzzy sets into a crisp output

Fuzzy set transformation to a crisp number is commonly called Defuzzification. It is a procedure that converts a Fuzzy set or Fuzzy number into a crisp value or number. The weighted average method (z^*) is the most popular used in fuzzy applications since it is one of the more computationally efficient methods and valid for symmetrical output membership functions. The algebraic expression gives z^*:

$$z^* = \frac{\sum_{i=1}^{n} \mu_{c_i}(x_i).(x_i)}{\sum_{i=1}^{n} \mu_{c_i}(x_i)} \quad (5)$$

Where c_1, c_2, ..., c_n are the output Fuzzy sets, and (x_i) is the value where the middle of each symmetric membership of the Fuzzy set c_i is observed.

Table 6: Fuzzy rule base of T, H, S and L

Rule No.	Condition	Action
1	**IF** $T_{current}$ = {L} **AND** T_{target} = {M} **THEN**	Heat
2	**IF** $T_{current}$ = {H} **AND** T_{target} = {M} **THEN**	Cool
3	**IF** $H_{current}$ = {L} **AND** H_{target} = {M} **THEN**	Fog
4	**IF** $H_{current}$ = {H} **AND** H_{target} = {M} **THEN**	Heat/Fan
5	**IF** $S_{current}$ = {L} **AND** S_{target} = {M} **THEN**	Water
6	**IF** $S_{current}$ = {H} **AND** S_{target} = {M} **THEN**	No water
7	**IF** $L_{current}$ = {L} **AND** L_{target} = {M} **THEN**	Lighting
8	**IF** $L_{current}$ = {H} **AND** L_{target} = {M} **THEN**	Curtain

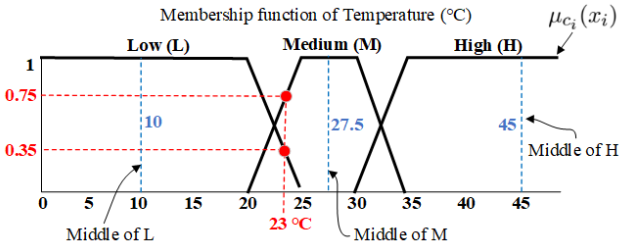


Figure 7: Fuzzy defuzzification

In Figure 7, if $\mu_{c_i}(x_i) = 1$, $x_{Low} = 10°C$, $x_{Medium} = 27.5°C$ and $x_{High} = 45°C$, we can estimate a crisp output of the temperature by using the z^* as follows:

$$z^* = \frac{(1).(10) + (1).(27.5) + (1).(45)}{(1+1+1))} = 27.5°C$$

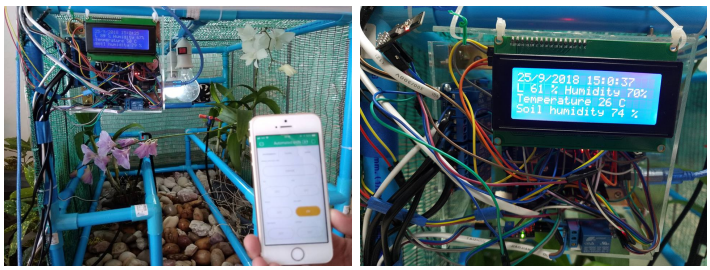
Also, the crisp output of humidity, soil moisture and sunlight: 65%, 65% and 60% respectively.

3.6 Control actuators

After the defuzzification process is executed successfully, the Fuzzy set value and crisp output will be used to control the actuators inside an orchid farm. For example, if the environment variables inside the orchid farm are now: $T_{current}$ = 21 °C (L = Low), $H_{current}$ = 54% (L), $S_{current}$ = 54% (L) and $L_{current}$ = 66% (H = High), the equipment (Actuators) inside the orchid farm must carry out the following tasks: Light bulbs = on (Heat), Fogs = on (Water), Sprinkles = on (Water) and Curtain = on (Sheltered greenhouse). The overview of the demonstration farm for orchid cultivation is shown in Figure 8. The methods of controlling the system can be done in two ways: allowing the system to operate automatically or manually via the Blynk software [21] on the mobile phone.

3.7 Loop control

The system is designed to work automatically, based on the knowledge that has been created from the previous steps, but users can still control the system by manual handling. The system reads various environmental data from the sensors in the fourth step and then uses such data to evaluate a Fuzzy value in the fourth step and a crisp value in the fifth step. The Fuzzy value will be calculated with the appropriate knowledge based rules (IF-THEN rules from the third step) for orchid planting. After that, the result will be taken to control actuators inside the smart orchid farming. The system works in a continuous loop without interruption.



(a) Controlling the system via a smartphone (b) Actuators inside the demonstration farm

Figure 8: Overview of demo orchid farming

3.8 Chat to Orchids (OrchidChatbot)

At the last phase (8), we design and create a chatbot to be used for talking to orchids to inquire about how healthy orchids are. The overview of chatbot's work is shown in Figure 9. According to Figure 9 (A), we use ChatterBot [22] running on Python version 3.7 to chat with orchids, which are grown inside the farm (D), through Blynk servers on the cloud (B), which act as intermediaries for communication. Communication between our chatbot (OrchidChatbot) and the Blynk servers to retrieve the environment information of the orchid farm, we use the Blynk library for Python [23]. OrchidChatbot starting from the smart farm control system (D) sends a Fuzzy value of the temperature (T), humidity (H), soil the moisture (S) and sunlight (L) to store in the Blynk servers by using Blynk libraries (C), these Fuzzy values must be evaluated by Fuzzy Logic already (C). Next, OrchidChatbot retrieves environment information (T, H, S and L) from Blynk serves for processing natural language like Figure 3 together with the Orchid corpus as shown in Figure 10. The format of the ChatterBot corpus explains in Table 7, and OrchidChatbot program, which is coded in Python, is shown in algorithm 1.

Referring to the chatbot program in algorithm 1, the first line loads Blynk libraries to a chatbot for communicating against Blynk servers on the cloud. Line number 2 is importing the Chatbot module from the chatterbot library to

build our chatbot program. The user authentication to Blynk servers by a key is shown in the line number 3. Creating virtual pins to send/receive the temperature, humidity, soil moisture and light between our chatbot and the Blynk servers shows in the line 4 to 7 respectively. Next, the line number 8 and 9 creates a chatbot and trainer object; and the line 10 points to the address of the Orchid corpus for training data. The chatbot retrieves the environment values from the Blynk servers from the line number 13 to 16. Last, line number 17 to 30 are used to respond between users and the chatbot; and when users want to leave the chat, type "bye" (line number 31 - 33).

```
categories:
- health
conversations:
- - How are you?[T:L]
  - I am feeling cold
- - How are you?[T:M]
  - I am feling warm.
- - How are you?[T:H]
  - I'm feeling hot.
- - How are you?[H:L]
  - I am feeling dry.
- - How are you?[H:M]
  - I am feeling cool, being good.
- - How are you?[H:H]
  - I am feeling wet.
- - How are you?[S:L]
  - I am feeling that my roots are dry.
- - How are you?[S:M]
  - I feel that my roots are fine.
- - How are you?[S:H]
  - I feel that my roots are too wet.
- - Can I help you?[T:L]
  - Yes, I want the warm weather.
- - Can I help you?[T:M]
  - No, it is already warm.
- - Can I help you?[T:H]
  - Yes, I want you to fog to reduce my heat.
- - Can I help you?[H:L]
  - Yes, I want more heat.
- - Can I help you?[H:M]
```

Figure 10: Examples of Orchid corpus

4 Performance evaluation

The performance evaluation in this research is divided into three components: 1) growth rate of orchid inflorescence length (cm./a week) planing inside the smart farming for ten weeks; 2) satisfaction from users who control and view the information via smartphones by the average ($\bar{x}$); and the accuracy of the conversation between the users, chatbot and the orchids by Dialogue specification and Harmonic mean, which the equations are shown as follows:

The average ($\bar{x}$):

$$\bar{x} = \frac{1}{n}.\sum_{i=1}^{n} x_i \tag{6}$$

Where $x_1, x_2, ..., x_n$ are the observed values of the sample testers, $\bar{x}$ is the mean value of these observations, and n is the number of observations in the sample. In this research, the number of testers (n) is thirty.

Algorithm 1 OrchidChatbot

```
1: import Blynk libraries
2: import ChatBot libraries
3: SET auth_token by 'AuthToken'
4: SET Temperature by Blynk_virtual_pin(1)
5: SET Humidity by Blynk_virtual_pin(2)
6: SET Soil_moisture by Blynk_virtual_pin(3)
7: SET Sunlight by Blynk_virtual_pin(4)
8: Create a chatbot object from ChatBot module
9: Create a trainer object from ChatterBotCorpusTrainer
10: SET the Orchid corpus path directory (namely orchid.yml)
11: Training Orchid corpus
12: while True:
13:     Temp ⟵ get temperature from Blynk server
14:     Humid ⟵ get Humidity from Blynk server
15:     Soil ⟵ get Soil_moisture from Blynk server
16:     Light ⟵ get Sunlight from Blynk server
17:     message = input('you:')
18:     if message.strip()!='bye':
19:         if Temp != NULL:
20:             reply = chatbot.get_response(message+Temp)
21:             print('ChatBot :', reply)
22:         if Humid != NULL:
23:             reply = chatbot.get_response(message+Humid)
24:             print('ChatBot :', reply)
25:         if Soil != NULL:
26:             reply = chatbot.get_response(message+Soil)
27:             print('ChatBot :', reply)
28:         if Light != NULL:
29:             reply = chatbot.get_response(message+Light)
30:             print('ChatBot :', reply)
31:     if message.strip()=='bye':
32:         print('ChatBot: bye')
33:         break
```

The Dialogue specification:

$$Dialogue\ specification = \frac{\sum_{j=1}^{m} wc_j}{\sum_{i=1}^{n} wa_i}; n >= m \tag{7}$$

Where wc_j are the words that chatbot answered correctly, and wa_i are all the words that chatbot answered.

The harmonic mean:

$$H = \left(\frac{\sum_{i=1}^{n} x_i^{-1}}{n}\right)^{-1} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \tag{8}$$

Where $x_1, x_2, ..., x_n$ are positive real numbers. For example, suppose that the user and chatbot chat together, as shown in Table 8, the results from the calculation with equation 7 and 8:

Table 8: Examples of conversation between the user and chatbot

User asked	Chatbot answered	Possible answers in Corpus
How are you?	I am cold.	I am feeling cold. I am feeling warm. I'm feeling hot.
Are you hot?	Yes, I am hot	No, I am so happy. Yes, I am very hot. No, I am so so.

The dialogue specification of question 1 and 2 are equal to 1.0 because the chatbot can answer both questions correctly. The harmonic mean of both questions:

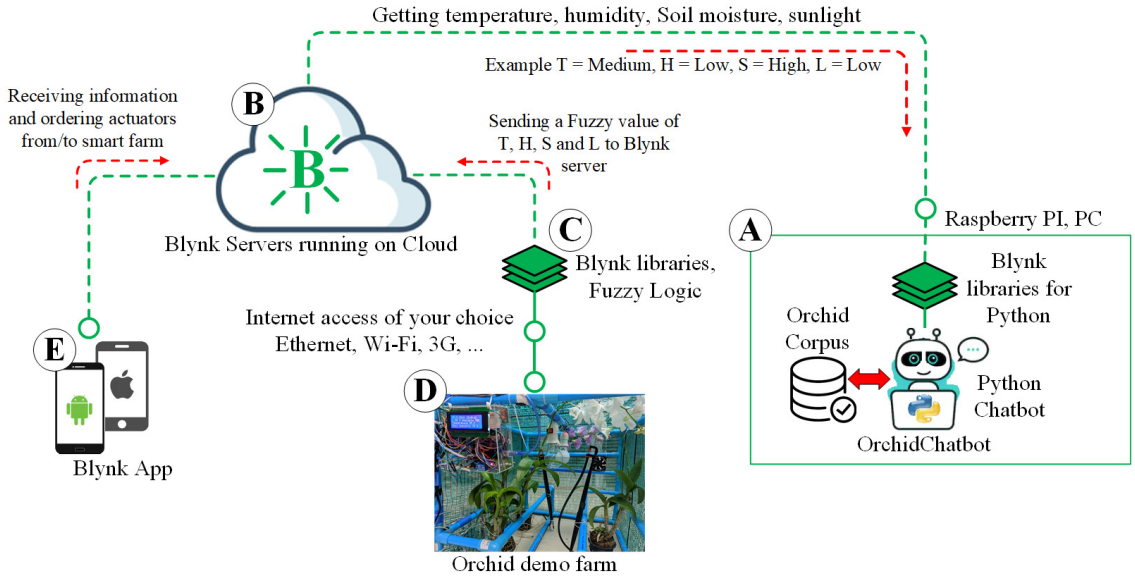


Figure 9: The working process of chatbot in Orchid demo farm

Table 7: The format of ChatterBot corpus

Syntax	Description	Example
categories:	Category of questions	categories: - health
conversations:	Start the conversation of each question category	conversations: - - How is your health? - I'm fine.
- - the question[?]	Questions that users ask to Chatbot	- - Hello - - How are you?
- the answer[.]	The answers which the Chatbot replies to the users	- Hi - I'm also good.
[X:Y]	Fuzzy value of each environment variable	- [T:L] - [H:M]

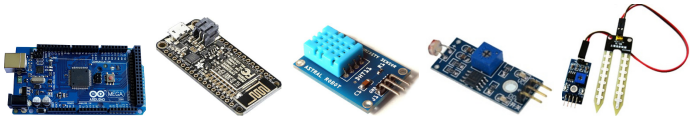
$$H = \frac{2}{\sum_{i=1}^{2} \frac{1}{x_i}} = \frac{2}{\frac{1}{x_1} + \frac{1}{x_2}} = \frac{2}{\frac{1}{1} + \frac{1}{1}} = 1.0$$

We also use the *precision* and *recall* methods, which are the classic techniques to evaluate a chatbot chatting. We create a total of 60 questions (Orchid corpus) in the discussion between users and orchids in this research listed in Table 10.

5 System development

5.1 Hardware for smart orchid farming

The devices such as sensors and actuators used in the orchid farm are as follows. Arduino Mega 2560 and Esp8266 module is used to control the entire system of automated smart farming. According to the sensors: the temperature and humidity (DHT11), soil moisture and LDR photo-resistor sensor as illustrated in Figure 11. The actuators are water pumps (12V DC), cooling fan (220V 50Hz), light bulbs (220V 50W), motors (12V DC), LED bulbs (220V 400 lumen), fogs and sprinkles as illustrated in Figure 12.



(a) Maga2560 (b) esp8266 (c) DHT11 (d) LDR (e) Soil

Figure 11: Arduino Mega, ESP8266 and sensors

(a) Light/LED (b) Fan (c) Pump (d) Sprinkler (e) Fog

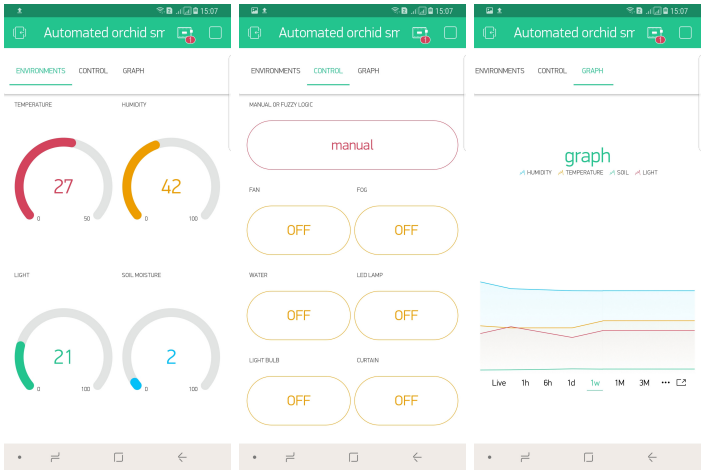
Figure 12: Actuators working in smart farm

5.2 Software for smart orchid farming

Orchid smart farming is controlled by a software developed from Blynk [21] application. It can be configured in two modes: the automatic and manual mode. According to the manual mode, orchid farmers can command the system via a smartphone by themselves. Another mode can operate automatically by itself. The sample interfaces of Blynk represent in Figure 13. For example, Figure 13 (a) displays the information in the Guage style, the operation buttons of various devices are shown in Figure 13(b) and Figure 13(c) shows usage statistics in the graph style.

5.3 OrchidChatBot hardware and software

The hardware and software on the chatbot side consist of the Linux operating system running on the Raspberry PI, Python 3.7, Blynk libraries for Python and ChatterBot for Python as illustrated in Figure 9 (A). The OrchidChatBot algorithm is explained in Algorithm 1.



(a) Gauge reports (b) Manual control (c) Graph reports

Figure 13: Blynk software for controlling the orchid farming

6 The experimental results

The results of the system evaluation are divided into three parts:

6.1 Results of system control via a mobile

Controlling smart orchid farming via mobile phones has been tested against 30 users. The results from the following tests shown in Table 9. Overall, the control system developed from Blynk has an average above 4.5 almost every item because the interfaces are developed to be exciting and easy to use.

Table 9: Satisfaction evaluation of users

Topic of evaluation	Satisfaction value ($\bar{x}$)
System design satisfaction	
1. Interesting system	4.8
2. Modernization of the system	4.7
3. Easy to use and easy to understand	4.9
4. System simulation is appropriate	4.4
Satisfaction in using the system	
1. Accuracy of the system	4.7
2. Appropriate to needs of users	4.6

6.2 Results of orchids planting inside the smart farming

The orchid farm is tested for a total of eight weeks from April to May 2018. We evaluate the growth of the inflorescence length of the orchids inside the farm every day. At night, the lighting system stops working to allow the orchids to get rid of the carbon dioxide. The growth statistics obtained from the experiments are shown in Figure 14. According to the line chart, the growth rate of the inflorescence length under a traditional farm is not stable because orchids proliferate

during the rainy season. The length of the inflorescence increases by an average of 34.3 centimeters (cm.) within ten weeks. In contrast, orchids are grown under an automatic smart farm; the growth rate is relatively constant at about 27.38 cm. only. Although in the overall, the growth rate of orchids inflorescence length raised in a traditional farm is higher than automated smart farm, the steady growth rate inside an automated smart farming can control the yield of orchids better than in practice. Since farmers can evaluate the production of orchids more precisely.

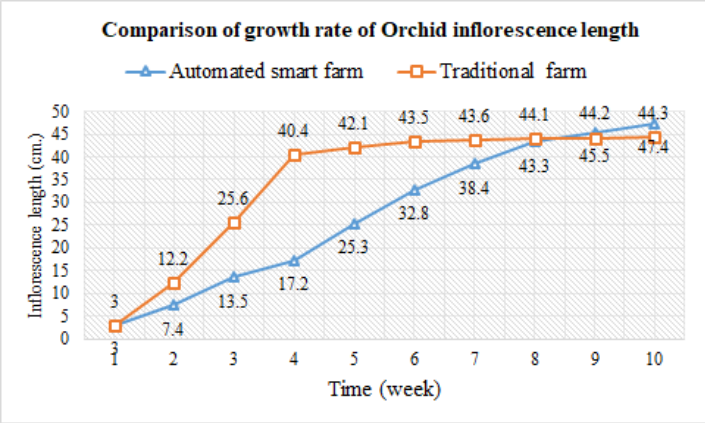


Figure 14: Comparison of the growth rate of the orchids inside the traditional farm against an automated smart farming

6.3 *Results of chatting between orchids and our chatbot*

The conversation between the orchids and chatbot is carried out through the command-line interface of Python shell, shown in Figure 15. Messages used to chat are related to the health of the orchids based on the information that the sensors can detect. Examples of questions and possible answers used in this research listed in Table 10.

The conversation is evaluated using the dialogue specification [24] and harmonic mean in the equation (7, 8). The average accuracy (Harmonic mean) of chatting between the user and the orchid is equal to 0.71, the *precision* = 0.75 and *recall* = 0.6 respectively.

7 Conclusion and future work

In this paper, we propose a chatbot which is used to chat with orchids planting inside the automated smart farming controlled by IoT and Fuzzy Logic, namely OrchidChatBot. The results of the system testing are divided into three parts: the average satisfaction of users using the mobile application to control an automated smart farming is equal to 4.68, the average growth rate of orchid inflorescence cultivated in the smart farm is 27.38 cm., and the average accuracy of the conversation between the chatbot and orchids are 0.71 (harmonic mean), 0.75 (precision) and 0.6 (recall) respectively. Besides, the significant advantages of automatic smart farming are reducing fertilizer, water, chemicals, insects, and yield control.

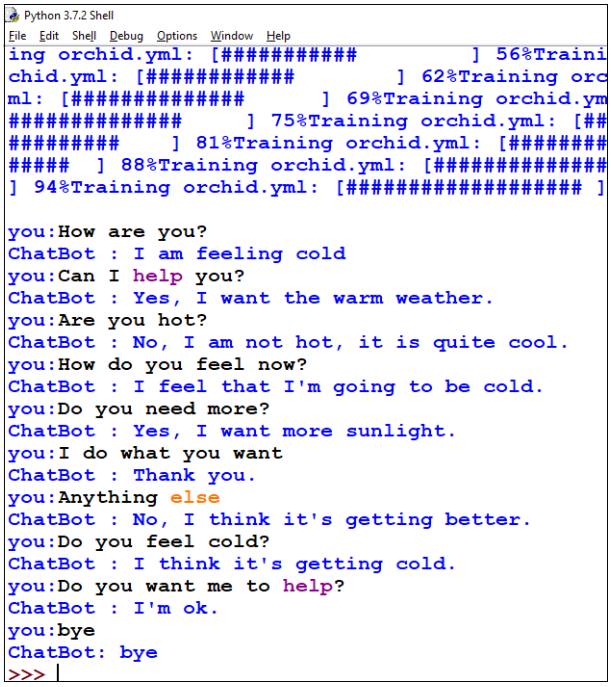


Figure 15: Chatting between Orchids and the chatbot by the command-line interface

Table 10: Examples of questions and possible answers

Question	Possible answers
How are you?	I am feeling cold. I am feeling warm. I'm feeling hot. I am feeling dry. I am feeling cool, being good. I am feeling wet. I am feeling that my roots are dry. I feel that my roots are fine. I feel that my roots are too wet.
Can I help you?	Yes, I want the warm weather. No, it is already warm. Yes, I want you to fog to reduce my heat. Yes, I want more heat. No, I am already invigorating. Yes, I want a blanket to increase warmth.
How do you feel now?	I feel that I'm going to be cold. I feel that I'm very comfortable. I feel that the weather is getting hotter.
Do you need more?	Yes, I want more sunlight. No, I do not more. Yes, I want to reduce the sunshine.
...	...
Are you hot?	No, I am so happy. Yes, I am very hot. No, I am not hot, it is quite cool. No, I am so so.

Moreover, chatting with orchids will help farmers to recognize the health of orchids, and to resolve problems that will occur with orchids more accurately.

In the future, we will enhance the conversation between the chatbot and orchids by using voice instead the command-line interface.

References

[1] Anindita Roy Chowdhury and Anshu Gupta, "Effect of Music on Plants – An Overview" International Journal of Integrative Sciences,

Innovation and Technology (IJIIT)., **4**(6), 30 – 34, 2015.

[2] Munasinghe, D. S. P, Weerakoon, S. R, Somaratne, S, "The effect of Buddhist pirith chanting and Western pop music on growth performance of "Pranajeewa", Codariocalyx motorius (Houtt.) H. Ohashi" Ceylon Journal of Science, 47(4), 357–361 2018, DOI: http://doi.org/10.4038/cjs.v47i4.7553

[3] F. H. Zakariya, M. Rivai and N. Aini, "Effect of automatic plant acoustic frequency technology (PAFT) on mustard pakcoy (Brassica rapa var. parachinensis) plant using temperature and humidity parameters," 2017 International Seminar on Intelligent Technology and Its Applications (ISITIA), Surabaya, 2017, pp. 334-339. doi: 10.1109/ISITIA.2017.8124105

[4] C. J. Baby, F. A. Khan and J. N. Swathi, "Home automation using IoT and a chatbot using natural language processing," 2017 Innovations in Power and Advanced Computing Technologies (i-PACT), Vellore, 2017, pp. 1-6. doi: 10.1109/IPACT.2017.8245185

[5] M. Muslih et al., "Developing Smart Workspace Based IOT with Artificial Intelligence Using Telegram Chatbot," 2018 International Conference on Computing, Engineering, and Design (ICCED), Bangkok, Thailand, 2018, pp. 230-234. doi: 10.1109/ICCED.2018.00052

[6] R. Angeline, T. Gaurav, P. Rampuriya and S. Dey, "Supermarket Automation with Chatbot and Face Recognition Using IoT and AI," 2018 3rd International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2018, pp. 1183-1186. doi: 10.1109/CESYS.2018.8723978

[7] B. Kohli, T. Choudhury, S. Sharma and P. Kumar, "A Platform for Human-Chatbot Interaction Using Python," 2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT), Bangalore, India, 2018, pp. 439-444. doi: 10.1109/ICGCIoT.2018.8753031

[8] C. E. Kaed, A. Ponnouradjane and D. Shah, "A Semantic Based Multi-Platform IoT Integration Approach from Sensors to Chatbots," 2018 Global Internet of Things Summit (GIoTS), Bilbao, 2018, pp. 1-6. doi: 10.1109/GIOTS.2018.8534520

[9] S. Khummanee, S. Wiangsamut, P. Sorntepa and C. Jaiboon, "Automated Smart Farming for Orchids with the Internet of Things and Fuzzy Logic," 2018 International Conference on Information Technology (InCIT), Khon Kaen, 2018, pp. 1-6. doi: 10.23919/INCIT.2018.8584881

[10] Chamovitz, D., What a Plant Knows: A Field Guide to the Senses, Scientific American, 2013.

[11] A. Saokaew, O. Chieochan and E. Boonchieng, "A smart photovoltaic system with Internet of Thing: A case study of the smart agricultural greenhouse," 2018 10th International Conference on Knowledge and Smart Technology (KST), Chiang Mai, 2018, pp. 225-230. doi: 10.1109/KST.2018.8426071

[12] F. A. Rafi, A. Kowanda, N. S. Salahuddin and T. Saptriani, "Design of Orchid Monitoring System Based on IoT," 2018 Third International Conference on Informatics and Computing (ICIC), Palembang, Indonesia, 2018, pp. 1-6. doi: 10.1109/IAC.2018.8780419

[13] N. Boonnam, J. Pitakphongmetha, S. Kajornkasirat, T. Horanont, D. Somkiadcharoen, J. Prapakornpilai, "Optimal Plant Growth in Smart Farm Hydroponics System using the Integration of Wireless Sensor Networks into Internet of Things", Advances in Science, Technology and Engineering Systems Journal, vol. 2, no. 3, pp. 1006-1012 (2017).

[14] A. Kamilaris, F. Gao, F. X. Prenafeta-Boldu and M. I. Ali, "Agri-IoT: A semantic framework for Internet of Things-enabled smart farming applications," 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT), Reston, VA, 2016, pp. 442-447. doi: 10.1109/WF-IoT.2016.7845467

[15] M. H. Memon, W. Kumar, A. Memon, B. S. Chowdhry, M. Aamir and P. Kumar, "Internet of Things (IoT) enabled smart animal farm," 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2016, pp. 2067-2072.

[16] R. C. Andrew, R. Malekian and D. C. Bogatinoska, "IoT solutions for precision agriculture," 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, 2018, pp. 0345-0349. doi: 10.23919/MIPRO.2018.8400066

[17] N. Putjaika, S. Phusae, A. Chen-Im, P. Phunchongharn and K. Akkarajitsakul, "A control system in an intelligent farming by using arduino technology," 2016 Fifth ICT International Student Project Conference (ICT-ISPC), Nakhon Pathom, 2016, pp. 53-56. doi: 10.1109/ICT-ISPC.2016.7519234

[18] S. Suakanto, V. J. L. Engel, M. Hutagalung and D. Angela, "Sensor networks data acquisition and task management for decision support of smart farming," 2016 International Conference on Information Technology Systems and Innovation (ICITSI), Bandung, 2016, pp. 1-5. doi: 10.1109/ICITSI.2016.7858233

[19] Melin P., Castillo O., Kacprzyk J., Reformat M., Melek W., Fuzzy Logic in Intelligent System Design, Springer International Publishing, 2018.

[20] Sudhir K. Mishra, Artificial Intelligence and Natural Language Processing, Cambridge Scholars Publisher, 2018.

[21] Blynk. (2018, 12 Oct). The mobile app for the IOT. [Online]. Available: https://www.blynk.cc/

[22] ChatterBot. (2019, 10 Aug). How ChatterBot Works. [Online]. Available: https://chatterbot.readthedocs.io/en/stable/

[23] Python client for Blynk. (2019, 10 Aug). Blynk library for Python 2, Python 3, and MicroPython. [Online]. Available: https://github.com/vshymanskyy/blynk-library-python

[24] S. Ghose and J. J. Barua, "Toward the implementation of a topic specific dialogue based natural language chatbot as an undergraduate advisor," 2013 International Conference on Informatics, Electronics and Vision (ICIEV), Dhaka, 2013, pp. 1-5. doi: 10.1109/ICIEV.2013.6572650

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 174-185 (2019)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

DFIG Defects Diagnosis Method for Wind Energy Conversion Chain

Fatima El Hammouchi[1,*], Lamia El Menzhi[2], Abdallah Saad[1]

[1]National Higher School of Electricity and Mechanic, Hassan II University, Casablanca 8118, Morocco

[2]National School of Applied Sciences, Abdelmalek Essaadi University, Tangier 9000, Morocco

ARTICLE INFO

Article history:
Received: 18 August, 2019
Accepted: 23 September, 2019
Online: 08 October, 2019

Keywords:
Diagnosis
Defect
Doubly fed induction generator
Frequency spectrum
Fast Fourier Transform
Lissajous curve
Wind turbine

ABSTRACT

This paper is an extension of research work originally presented in 2018 IEEE fifth International Congress on Information Science and Technology (CiSt). The research consists on developing method to diagnose electrical defects affecting wind turbine doubly-fed induction generator DFIG which constitutes a crucial part of wind energy conversion chain. First off all, we create a model of a non-defected wind conversion system based on mathematical equations introduced in Matlab Simulink. Then, we apply an indirect vector control stator field orientation in order to increase wind energy performance. With the aim of diagnosing the defects attacking wind turbine generator, we propose a method based on grouping of fast Fourier transform spectral analysis and Lissajous curves performed to generator stator and rotor currents. This diagnosis technique is applied to wind turbine in normal operation (non-defected generator) in order to have a reliable reference data for asynchronous generator behaviour. However, connected to the grid, wind turbine generator is affected by various faults occurring in electrical power networks. Therefore, the diagnosis method is applied also to a defected generator. Considering diversity of grid defects, we deal in the current paper with open stator supplying phases and open rotor feeding phases due to rotor side converter legs opening. Indeed, this diagnosis method allows diagnosing generator defects type and severity by comparing the resulting frequency spectrum analysis and Lissajous curves under abnormal condition operating to reference data obtained in case of non-defected generator. So, our proposed method contributes to DFIG defects identification and anticipation. The simulations had been accomplished using Matlab Simulink. These results proved the efficiency and effectiveness of the proposed DFIG diagnosis method for wind energy conversion chain.

1. Introduction

To satisfy the world increasing energy demand we need to add new capacity to the grid while reducing global CO2 emissions and conserving our environment. Without doubt renewable energy in general and wind power in particular can form pillar supporting economic development and growing energy needs.

Indeed, the 51.3 Gigawatts of new installations bring total cumulative installations up to 591 Gigawatts by the end of 2018 [1]. This worldwide capacity showed that wind sector continues to progress, draining investments in wind energy.

Morocco take advantage of the latest technology in wind field and increase installed wind power to 2 Gigawatts by 2020 [2]. Actually, the overwhelming majority of Moroccan wind farms house horizontal axis wind turbines. For these wind turbines variable speed industry, the doubly fed induction generator DFIG is the most used [3]. Thanks to its excellent operational and control features [4], the DFIG makes wind energy integration with electrical networks easy and effective. In fact, much research focused on how to improve the integration of wind energy in the grid. This problematic have double face.

On one hand, wind power is intermittent energy source. So, integration of the wind energy can unfortunately affect the power system negatively. Furthermore, the electrical grids have been conventionally conceived for unidirectional energy flows from power stations going to cities. The use of dispersed wind energy

* Fatima EL HAMMOUCHI, National Higher School of Electricity and Mechanic, Hassan II University, Casablanca 8118, Morocco, Contact: +212 668 085 468, Email: elhammouchi.fatima@gmail.com

generators is more likely to result in bi-directional flow and may exacerbate problems with voltage and fault management [5]. Hence, in order to inject wind power to the grid, wind turbines generators must have the ability to contribute to both the voltage and frequency control in stabilising the power system following disturbance [6].
On the other hand, wind turbines generators must be connected to reliable distribution networks. So, grid operators need to ensure that electric networks operate continuously in a safe way [7]. But, faults in electrical power networks are inevitable and unpredictable events. As, wind turbines are connected to grid, they are consequently exposed to various grid faults. Unfortunately, faults in electrical grids can immediately paralyze electricity production in wind farms and sometimes lead to a loss of total structure. Actually, when grid fault occurs wind turbine is disconnected in order protect the structure and reconnected when normal operation has been resumed. However, for large wind energy capacity disconnection from the grid could generate problems in the control of frequency and voltage in the system [8]. Thus, wind turbines generators must be able to avoid excess fault levels while still contributing to fault identification and stop their propagation to save money and time. For this reason, it is necessary to know behaviours of DFIG wind turbine when grid faults are detected.

Consequently, the effectiveness DFIG defects diagnosis remains a necessity.

In the current paper a diagnosis method is developed for wind turbine doubly-fed induction generator. This method is based on analysis of frequency spectrums and Lissajous curves of DFIG stator and rotor currents.

Firstly, we present the wind energy chain modeling based on mathematical equations created in Matlab Simulink.

Then, in order to increase performance of wind system, we chose to control stator flux by the application of an indirect field oriented control (IFOC) to DFIG. Afterwards, we use stator and rotor currents to establish frequency spectrums analysis and to draw Lissajous curves. This diagnosis method was applied in the first phase to a non-defected generator in order to have a reliable reference data in healthy case of DFIG. The second phase consists on using the method while generator is under most known abnormal condition operating due essentially to grid faults.

In the last part, the resulting Lissajous curves and frequency spectrums in case of defected generator are compared to Lissajous curves and frequency spectrums obtained when generator is under normal operating conditions (reference data: presented in previous article).

This method allows diagnosing wind turbine generator defects type and severity; thereafter, anticipating serious damage affecting wind energy conversion chain.

2. Wind Energy Chain Conversion Modeling

2.1. Wind Turbine Modeling

The essential components of wind turbine installed in Moroccan parks are presented in the figure bellow. The modeling is about horizontal-axis wind turbines HAWT using DFIG. In fact, the wind turbine generator is asynchronous machine doubly-fed: the stator and rotor are both connected to power supply. The DFIG stator is connected to national grid through protection contactor. The rotor is connected to the grid using the transformer and converters "back-to-back" composed of rotor side converter (RSC) and grid side converter (GSC). A capacitor is placed between two converters to reduce the voltage ripples. Indeed, regardless of the rotor power magnitude and direction, the GSC has to keep the dc-link capacitor voltage at a set value and to guarantee a converter operation with unity power factor (zero reactive power) [9].

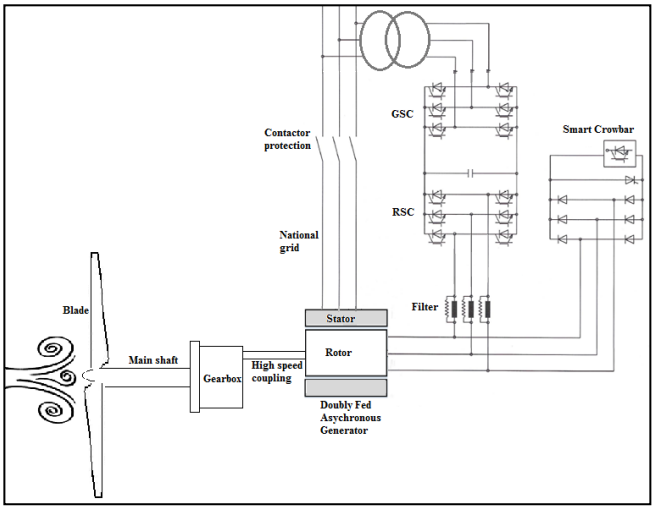


Figure 1: Main Components of Wind Turbine.

We introduced sum of different sinusoidal signals: varied amplitudes and frequencies in order to create wind speed model in Matlab Simulink as showed in the figure 2, which reflects approximately a recorded wind speed in Moroccan wind farms [10].

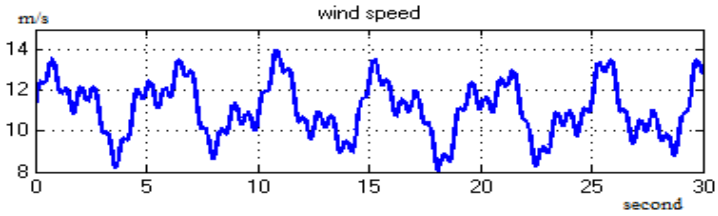


Figure 2: Wind Speed Model (m/s).

We assume that frictions are neglected and with the use of equations presented in previous paper [11], the wind turbine Matlab Simulink model is created.

Also, the DFIG modeling is based on equations of the stator and the rotor voltages (Us, Ur) in Park reference (d, q) as presented below:

$$\begin{cases} U_{ds} = \dfrac{d\varphi_{ds}}{dt} + R_s.i_{ds} - \varphi_{qs}\omega_s \\ U_{qs} = \dfrac{d\varphi_{qs}}{dt} + R_s.i_{qs} + \varphi_{ds}\omega_s \\ U_{dr} = \dfrac{d\varphi_{dr}}{dt} + R_r.i_{dr} - \varphi_{qr}(\omega_s - \omega_r) \\ U_{qr} = \dfrac{d\varphi_{qr}}{dt} + R_r.i_{qr} + \varphi_{dr}(\omega_s - \omega_r) \end{cases} \quad (1)$$

With:

$$\omega_r = \omega_s - P.w \quad (2)$$

In park reference, the flux equations can be expressed as follows:

$$\begin{cases} \varphi_{ds} = L_s.i_{ds} + M.i_{dr} \\ \varphi_{qs} = L_s.i_{qs} + M.i_{qr} \\ \varphi_{dr} = L_r.i_{dr} + M.i_{ds} \\ \varphi_{qr} = L_r.i_{qr} + M.i_{qs} \end{cases} \quad (3)$$

$$L_s = l_s - M_s, \qquad L_r = l_r - M_r$$

The electromagnetic torque is expressed according to equation below:

$$T_{em} = p(\varphi_{ds}.i_{qs} - \varphi_{qs}.i_{ds}) \quad (4)$$

Where **p** represents pole pairs number of the DFIG.

We chose to control DFIG by stator flux. We seek to cancel the indirect stator flux in order to keep that direct stator flux. In this mode of DFIG operating conditions, control of the torque and flux is decoupled. This method is applied to improve the DFIG active and reactive power control performance, which allows controlling flux and torque dynamically and separately.

Based on equations (3) and (4) we have:

$$\begin{cases} i_{ds} = \dfrac{1}{Ls}(\varphi_s - M.i_{dr}) \\ i_{qs} = -\dfrac{M}{Ls}.i_{qr} \\ T_{em} = p\varphi_s.i_{qs} = -p\dfrac{M}{Ls}\varphi_s.i_{qr} \end{cases} \quad (5)$$

In general, for large wind turbine, the generator stator resistance is negligible.

Also, it is supposed that the electrical network is stable and the field is constant. Thus, by using equation (1) we get:

$$\begin{cases} U_{ds} = 0 \\ U_{qs} = \varphi_s\omega_s \end{cases} \quad (6)$$

So, the DFIG active and reactive powers of can be written as equations below:

$$\begin{cases} P_s = U_{ds}.i_{ds} + U_{qs}.i_{qs} \\ Q_s = U_{qs}.i_{ds} - U_{ds}.i_{qs} \end{cases} \quad (7)$$

We get the expression of the active and reactive powers:

$$\begin{cases} P_s = -U_s\dfrac{M}{Ls}.i_{qr} \\ Q_s = \dfrac{U^2{}_s}{\omega_s.L_s} - U_s\dfrac{M}{Ls}.i_{dr} \end{cases} \quad (8)$$

So, the equation (8) shows that the control of the active power is independent of the reactive power. Indeed, the active power is controlled by iqr the quadrature rotor current whereas the reactive power can be controlled by idr the direct rotor current.

$$\begin{cases} U_{dr} = (R_r + S(L_r - \dfrac{M^2}{Ls})).i_{dr} - g\omega_s(L_r - \dfrac{M^2}{Ls}).i_{qr} \\ U_{qr} = (R_r + S(L_r - \dfrac{M^2}{Ls})).i_{qr} + g\omega_s(L_r - \dfrac{M^2}{Ls}).i_{qr} + g\dfrac{M}{Ls}.U_s \end{cases} \quad (9)$$

Also, to perform the machine control power there are two approaches for the field oriented control direct and indirect:

- Direct field oriented control is simple to implement : it consist on ignoring the terms of coupling in equation (9) and to establish an independent regulator in each axis to control the active and reactive power independently. But the direct field oriented control is not the most efficient [12];
- Indirect field oriented control: in this control we consider the coupling terms in equation (9) and we look for compensating them to control the powers and rotor currents. The indirect field oriented control is being more frequently used as it guaranties the easy operation over all the speeds range [13].

Table 1: Symbols and Abbreviations for generator Modeling

Symbol	Quantity
φ_s **(d, q)**	Stator flux components in Park reference
φ_r **(d, q)**	Rotor flux components in Park reference
Rs	Stator resistance
Rr	Rotor resistance
Ls	Stator cyclic inductances
Lr	Rotor cyclic inductances
ls	Stator leakage inductances
lr	Rotor Leakage inductances
Ms, Mr	Mutual inductances between stator and rotor phases respectively
M	Maximum mutual inductance between stator and rotor stage (the axes of the two phases coincide)
i_s ***(d,q)***	DFIG stator current components in Park reference
i_r ***(d,q)***	DFIG rotor current components in Park reference
Tem	Electromagnetic torque
U_s ***(d,q)***	Stator voltage components in Park reference
Ur ***(d,q)***	Rotor voltage components in Park reference
P_s	DFIG active power
Q_s	DFIG reactive power

For DFIG modeling, we chose in this study to orient the stator flux of generator according to the direct axe. Moreover, we use the indirect vector field oriented control (IFOC) to increase the system performance. The wind energy conversion chain model created in Matlab Simulink is illustrated in figure 3 below.

The figure 3 presents the model of wind turbine created in Matlab Simulink which is composed of:

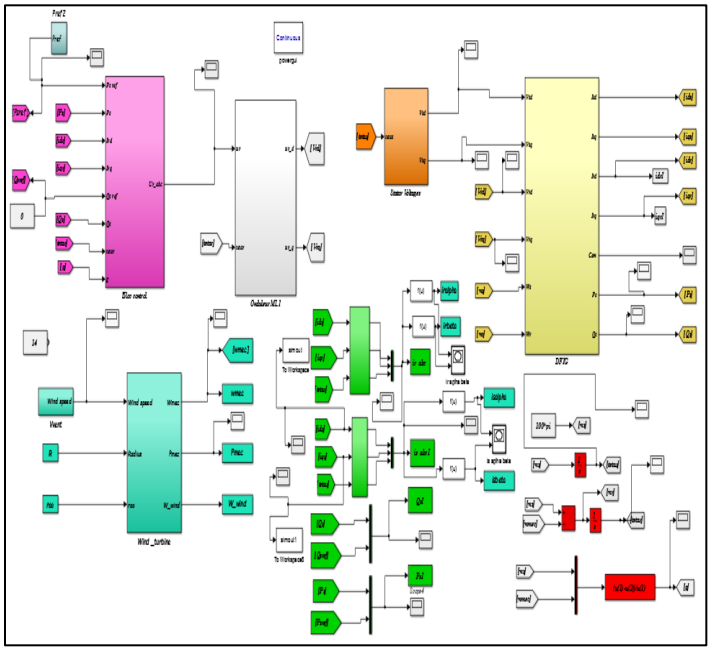


Figure 3: Wind Turbine Model in Matlab Simulink.

- The blue bloc at the bottom in the left contains modeling equations of wind turbine. In order to have maximum C_p we take $\beta = 0$. Also, the wind speed model introduced as illustrated in figure 2;
- The yellow bloc at the top on the left includes the DFIG stator and rotor flux and currents equations. This bloc is linked to the power supply (orange bloc);
- The chosen control of generator is implemented in the pink bloc at the top on the right. Actually, the DFIG indirect field oriented control allows us to control dynamically and separately flux and torque. The bloc is connected to the grey bloc of Pulse Width Modulation PWM;
- We use the Concordia transformation in order to obtain DFIG currents Lissajous curves as figured in green blocs;
- We use mathematic equations in red blocs in order to get the DFIG slip g.

2.2. Simulations and Results

The graphs below represent the simulation at the first thirty seconds of the wind turbine model created in Matlab Simulink. We suppose that frictions are neglected.

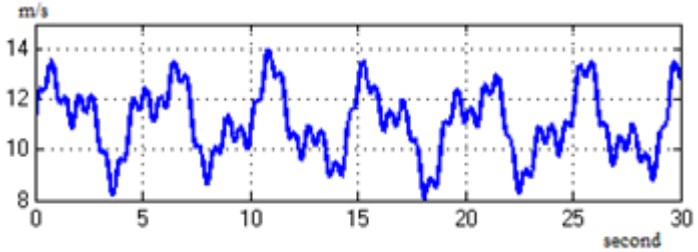


Figure 4: Wind Speed (m/s).

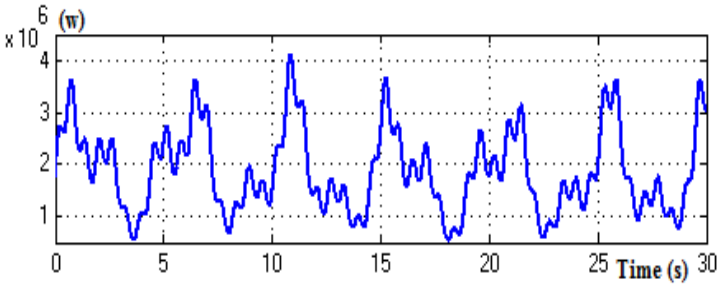


Figure 5: Power Captured by Blade (w).

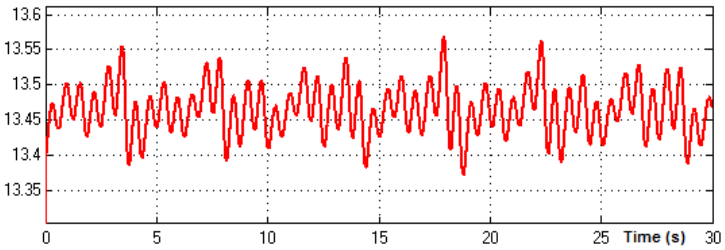


Figure 6: Specific Speed λ.

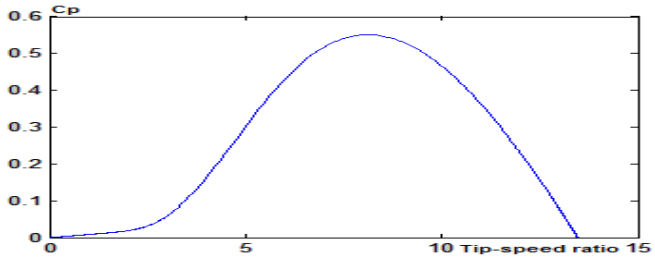


Figure 7: Cp Coefficient of Performance at β =0.

In fact, figure 4 represents the variable wind speed which we apply to wind system conversion. The figures 5 and 6 show the system performance. Actually, the power captured by blade follows the same pace of wind speed.

The figure 7 shows coefficient of performance as a function of specific speed λ and the angle of orientation of the blade β. The coefficient reaches its maximum value which is 0.56 when β is zero and λ is optimal. Using the indirect vector field-oriented control (IFOC), we got the following figures representing asynchronous generator performance.

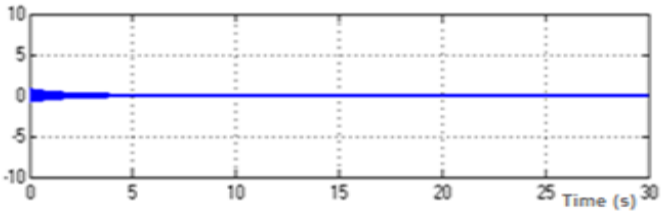


Figure 8: DFIG Indirect Stator Field φ_{qs} (wb).

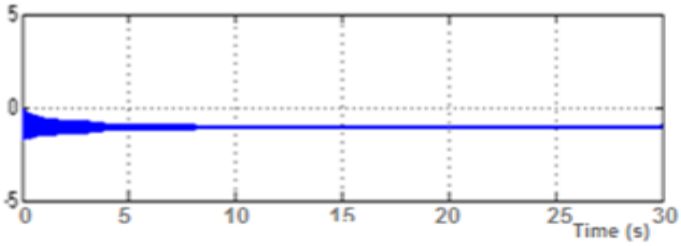


Figure 9: DFIG Direct Stator Field φ_{ds} wb.

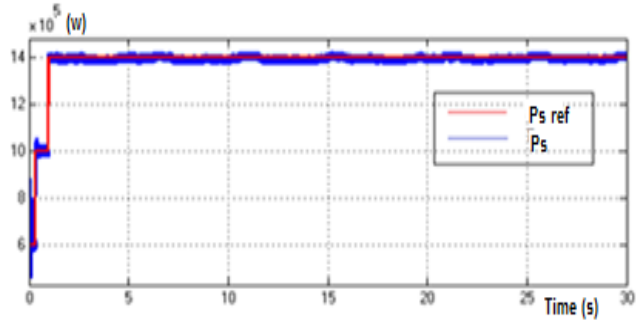


Figure 10: Comparison between Ps and Psref.

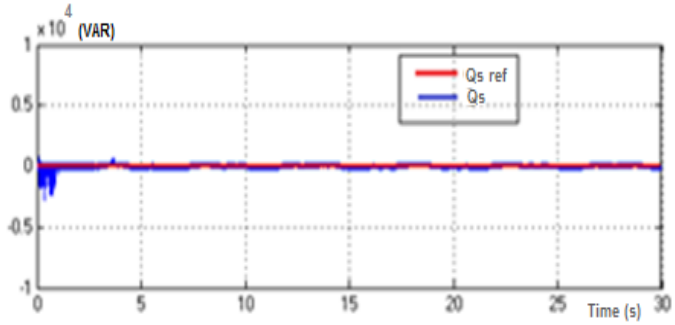


Figure 11: Comparison between Qs and Qsref.

In figures 8 and 9 the indirect field values of stator machine reached to zero and the direct stator field is constant since the stator flux of DFIG is chosen to be oriented following the direct axe. Furthermore, we choose Psref a reference for active power as a three steps form applied in different time. As illustrated in figure 10, the real active power pursues Psref. In addition, as showed in figure 11, the system reactive power follows Qsref which around zero value.

A zoom is presented in figure 12 of doubly-fed asynchronous generator three-phase stator currents.

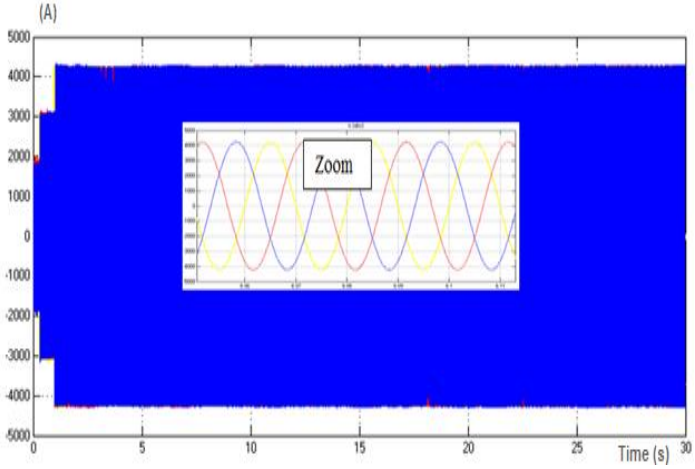
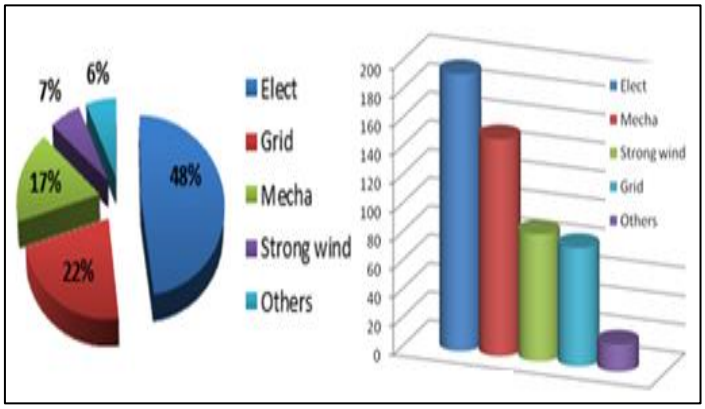


Figure 12: DFIG Three Phase Stator Currents.

It is shown that DFIG three-phase stator currents isabc have exactly sinusoidal shapes.

3. DFIG Diagnosis Defects Method

Nowadays, the electricity generated from wind turbines keeps increasing. However, faults in electrical power networks are inevitable events. According to data collected from yearly defects study of wind turbines Moroccan park [14], wind turbine connected to grid is exposed to various grid faults: 22% of defects are due to grid faults as presented in figure 13 (a). Also, grid defects can cause significant losses to wind system even if grid defects frequency remains less than other defects frequencies figure 13 (b) .



(a) (b)

Figure 13: Wind Turbines Defects: (a) type and (b) frequency.

The figure 13 shows defects type and frequency according to yearly Moroccan park data defects inspection.

Indeed, as indicated in sector graph figure 13.a wind turbines suffer from various internal electrical faults (48%), we mentioned for example over-current in DFIG stator phases, overvoltage in crowbar, overcurrent phase in GSC, thermal tripping in yaw drive motor because of cable isolation failure, etc . Also, linked to the electrical network, wind turbines are exposed to several grid faults (22%) such as lack of grid voltage, variation in the voltage magnitude and time duration and overheated generator due to the short circuit in the power supply, etc. The third defects type which represents 17% of total defects types is mechanical defects that occurs mainly in gearbox, electrical generator bearings, yaw motor, and also blades bearings. The rest defects types are essentially due to strong wind and other types.

The histogram above in figure 13.b shows different defects frequencies. It proves that the electrical defects are the most recurrent: they occur more than 194 times during a year in this park. In the second place come the mechanical defects with 152 times, followed by strong wind (89 times), grid faults (83 times) and others less than 20 times.

Usually, when fault occurs in the grid wind turbine is disconnected to avoid loss structure; however, for large wind energy capacity disconnection from the grid could generate problems. For this reason, multitudes of techniques are used to monitor the wind energy conversion chain health continually [15, 16].

In this current paper we propose diagnosis method for wind turbine DFIG electrical defects. The method is based on association of Fast Fourier Transform introduced and Lissajous curves analysis.

Indeed, Fast Fourier Transform (FFT) allows a conversion of a signal from its original domain to a representation in the frequency domain. The FFT analysis is applied in this paper to generator

stator currents in order to determine the harmonic components amplitude and to get the maximum amplitude spectrum of the wavelet coefficients. The resulting spectrum offers a source of defects information [17].

To FFT analysis we associate Lissajous curves analysis. Currently, in the field of Lissajous curves, recent developments are realized in different domains such as the rotating machinery field [18], and defects diagnosis [19]. The method to get Lissajous curves of stator and rotor currents was explained in detail in previous paper [20]. Thanks to the curves obtained we will have an idea about defects kind impacting generator in wind conversion energy chain. In fact, when the defects appear, the Lissajous curves will be deformed depending on defects kind.

In the next part diagnosis method will be applied firstly to a non-defected generator.

4. Diagnosis Method Applied in Case of DFIG Normal Condition Operating

In order to have a reliable reference data for behaviour of DFIG in healthy case, the diagnosis method is developed for generator during defects-free operation.
As showed in figure 14, the FFT diagram of the stator three-phase current is implemented in Matlab Simulink.

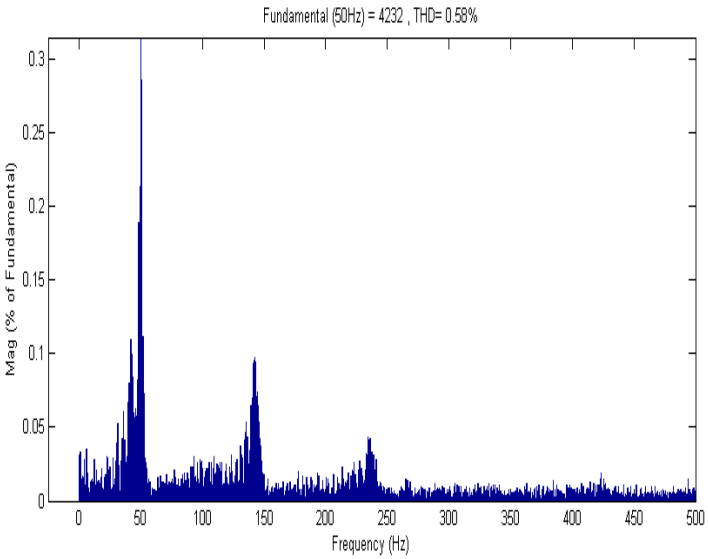


Figure 14: Fast Fourier Transform Spectrum of Stator Currents.

The signal fundamental frequency is at 50 Hz and the Total Harmonic Distortion (THD) which measures harmonics content is equal to 0.58%. In addition, harmonics are detected at frequency equals to 150 Hz and at 250 Hz. These harmonics are due to the use of PWM Pulse Width Modulation in the system, they represent the third and fifth harmonics. Also, magnitude percentages are respectively 0.1% and 0.05%.

Also, Lissajous curves of DFIG stator and rotor three-phase currents are simulated in Matlab Simulink.

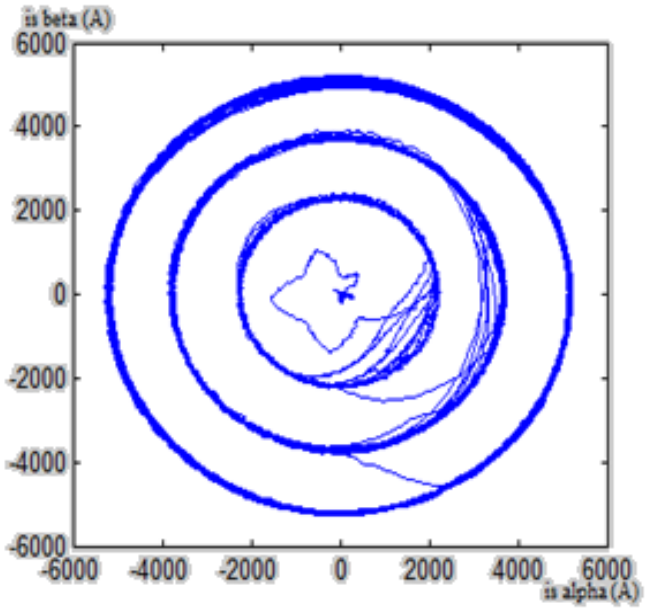


Figure 15: DFIG Stator Currents Lissajous Curves.

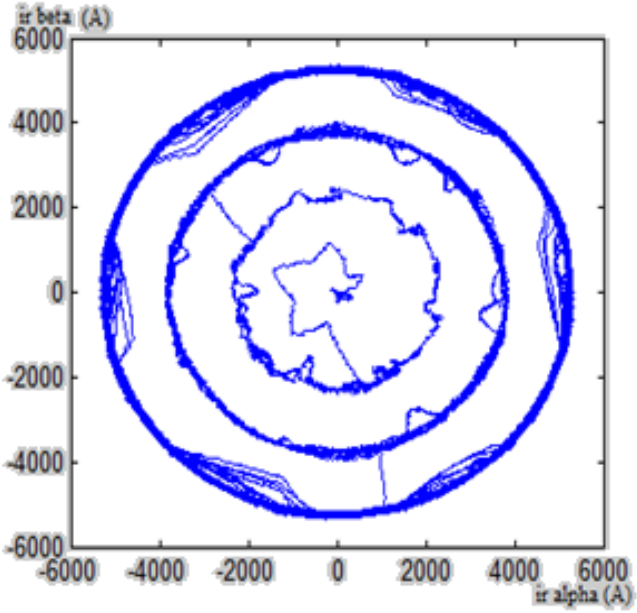


Figure 16: DFIG Rotor Currents Lissajous Curves.

The figures 15 and 16 illustrate DFIG stator and rotor currents Lissajous curves when generator is under normal operating condition. In this case, Lissajous curves have the form of three circles. These circles are the result of reference active power Psref which had been chosen as a signal of three steps presented in figure 10. Also, the rotor currents Lissajous curves have six petals flower shape (figure 16). This shape is explained by the fact that the rotor generator power supply is realized through the inverter with 6 switches.
In the next part we apply diagnosis method to defected DFIG.

5. Diagnosis Method Applied in Case of Open Stator Feeding Phases

As mentioned previously, larger wind turbines were developed and installed in windy world parts. Currently, the wind power

generated is comparable to conventional fossil-fuel energy.
So, the impact of grid faults on wind power plants is considerable and could lead to blackouts if we allowed disconnection of wind turbines from grid. To keep wind turbines connected to the grid despite the presence of faults, we need to monitor and diagnose continuously the structure. Indeed, wind turbine generator is subjects to grid faults [21, 22]. We mentioned lack or variation of the voltage grid magnitude and frequency, and short circuit. Some recurrent grid faults are diagnosed by using the proposed diagnosis method [23].
Considering diversity of grid defects, we deal in this part with the problem of open stator feeding phases. This defect occurs rarely. However, when it occurs, it mainly affects generator and can cause significant losses.
In fact, when the stator is fed from two phases, the generator will continue to operate but the phase imbalance will cause its heating which can lead to eventual damage to DFIG, thereafter the loss of the whole structure.

5.1. Open stator supplying phase 'a'

The figure 17 represents FFT of stator currents in case of generator fed from two phase power supply (phases 'b' and 'c').

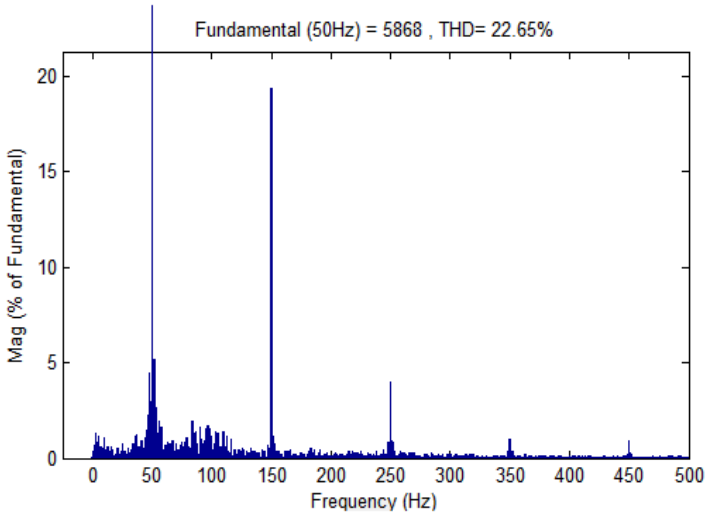


Figure 17: Stator Currents FFT Spectrum in Case of Open Stator Supplying Phase 'a'.

In addition to harmonics at frequency equals to 150 Hz and at 250 Hz (case of defects-free generator), peaks appear in new frequency (at 25 Hz and 75 Hz). Also, Total Harmonic Distortion THD passed from 0.58% to 22.65% in case of defected generator. Indeed, higher THD means increase of generator heating and peak currents. As showed in figure 9 fundamental magnitude was 0.32%. But, when one feeding stator phase 'a' was opened, this fundamental magnitude value increases and passed to 24%.

The following figures show Lissajous curves respectively generator stator and rotor currents in case of open stator supplying phase 'a'.

Comparing stator and rotor Lissajous curves obtained in cases of defects- free and defected generator, we see clearly in figures 19 and 20 that stator and rotor curves are deformed and the currents values increase.

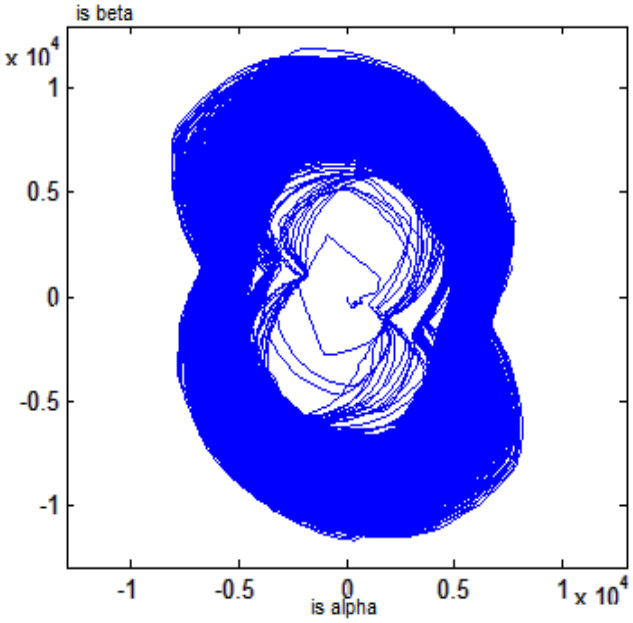


Figure 18: DFIG Stator Currents Lissajous Curves in Case of Open Stator Supplying Phase 'a'.

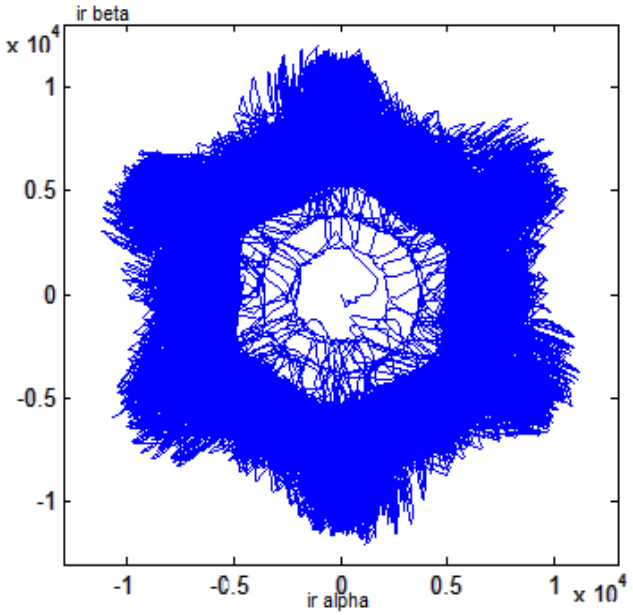


Figure 19: DFIG Rotor Currents Lissajous Curves in Case of Open Stator Supplying Phase 'a'.

In fact, the opening of one stator feeding phase (phase 'a') generates an increase of the currents in the other phases. Also, the stator currents beta reach higher values than alpha currents. The curves have now elliptical shapes according to the vertical axis (stator currents beta). Also, curves circumference and thickness increased, which makes difficult observation of three separate curves.

The same for rotor currents Lissajous curves showed in figure 19 they are superimposed on each other because of increasing curves thickness. Which makes impossible detection of the three separate curves as Lissajous curves obtained in case of normal condition

operating. Also, curves have six petals flower shape but they are deformed comparing to the reference curves.

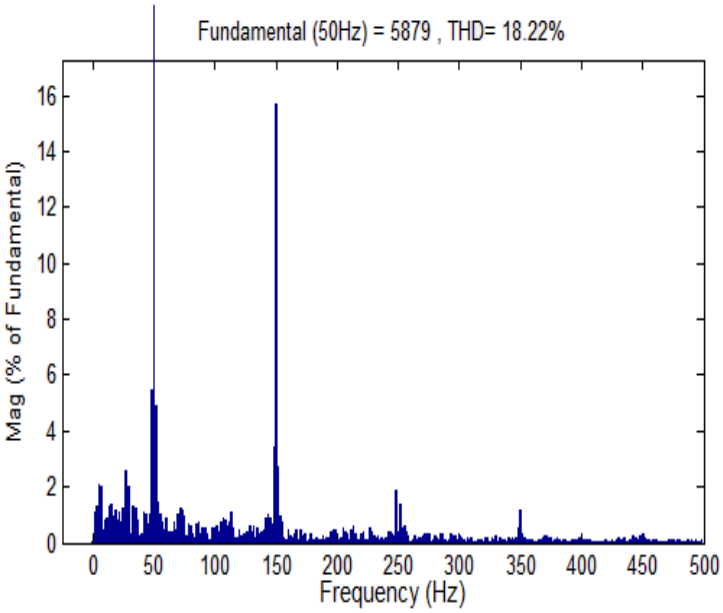


Figure 20: Stator Currents FFT Spectrum in Case of Open Stator Supplying Phase 'b'.

5.2. Open stator supplying phase 'b'

As illustrated in figure 20, the Fast Fourier Transform spectrum of DFIG stator currents when supplying phase 'b' is opened has approximately the same magnitude of fundamental in case of opened phase 'a'. Also, harmonics at frequency 150 Hz, 250 Hz and 350 Hz appear in the FFT spectrum.

In term of Total Harmonic Distortion THD in this case, the value is 18.22%, a little less comparing to open supplying phase 'a'.

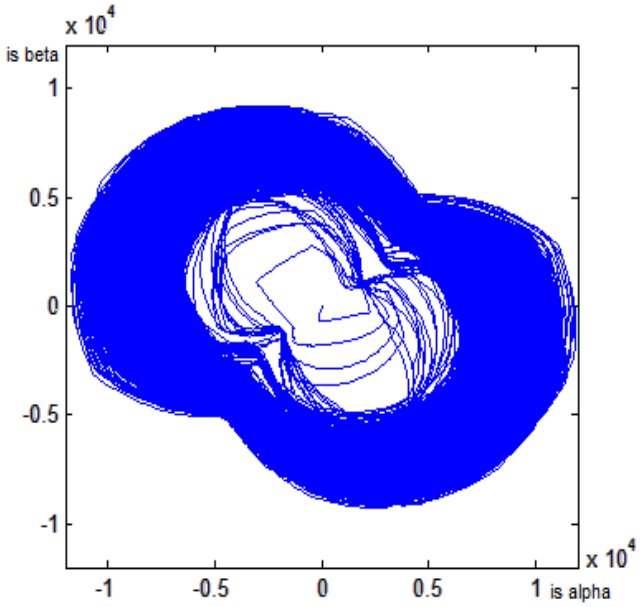


Figure 21: DFIG Stator Currents Lissajous Curves in Case of Open Stator Supplying Phase 'b'.

In case of open stator feeding phase 'b', stator and rotor Lissajous curves obtained respectively in figures 21 and 22, are deformed comparing to the case of opening phase 'a'.

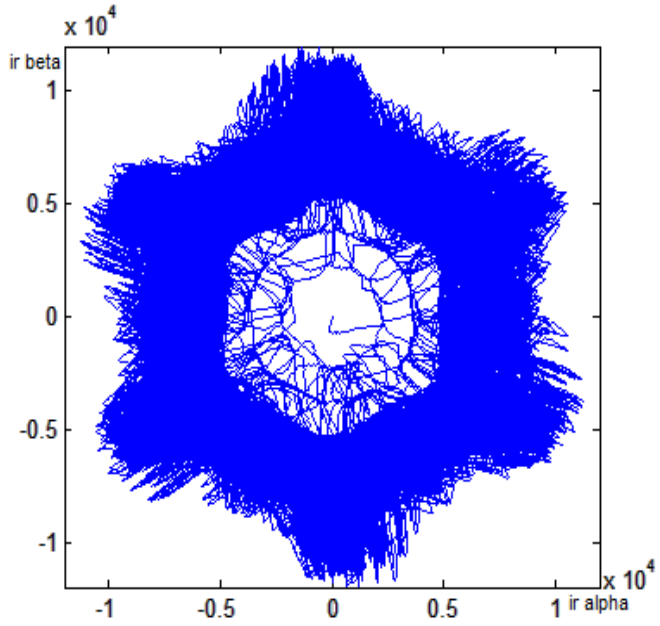


Figure 22: DFIG Rotor Currents Lissajous Curves in Case of Open Stator Supplying Phase 'b'.

Indeed, the stator currents curves have elliptical shapes. But, for open phase 'b' stator currents alpha reach higher values than stator currents beta. Also, we note an inclination angle of $2\pi/3$ of Lissajous curves compared to open phase 'a'.

For rotor currents Lissajous curves showed in figure 22, we note that the curves have the same shape as figure 19 case of open phase 'a' with a slight increase in rotor currents.

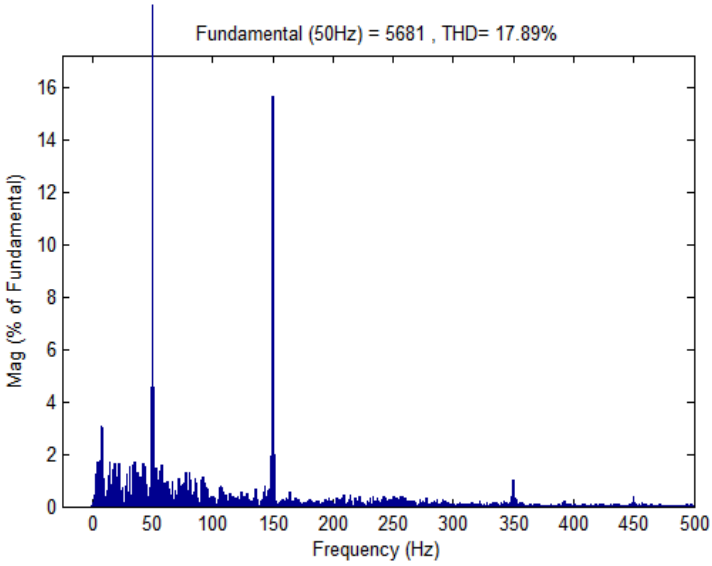


Figure 23: FFT Spectrum of Stator Currents Case of Open Stator Supplying Phase 'c'.

5.3. Open stator supplying phase 'c'

As showed in figure 23, the FFT spectrum of stator currents in the case of open supplying phase 'c' has approximately the same magnitude of fundamental in case of open phase 'a' or phase 'b'. In addition, the THD in the case of open supplying phase 'c' is 17.89 %, it has approximately the same value comparing to open supplying phase 'b'.

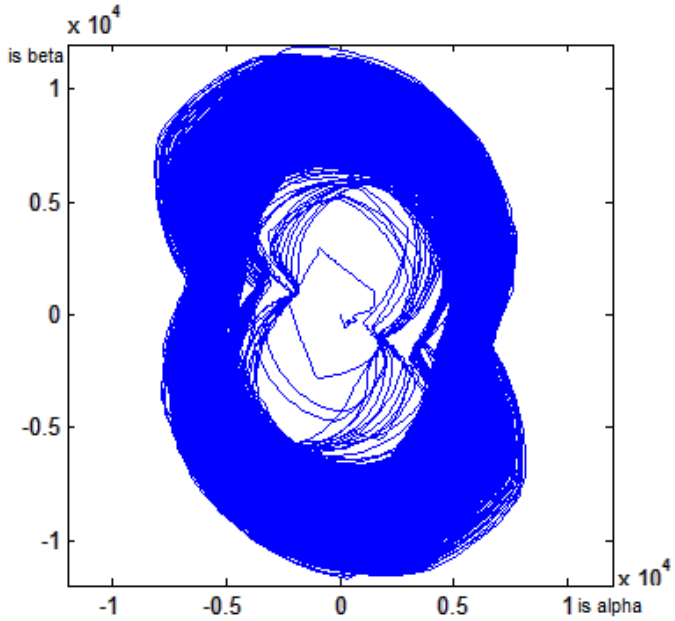


Figure 24: DFIG Stator Currents Lissajous Curves in Case of Open Stator Supplying Phase c.

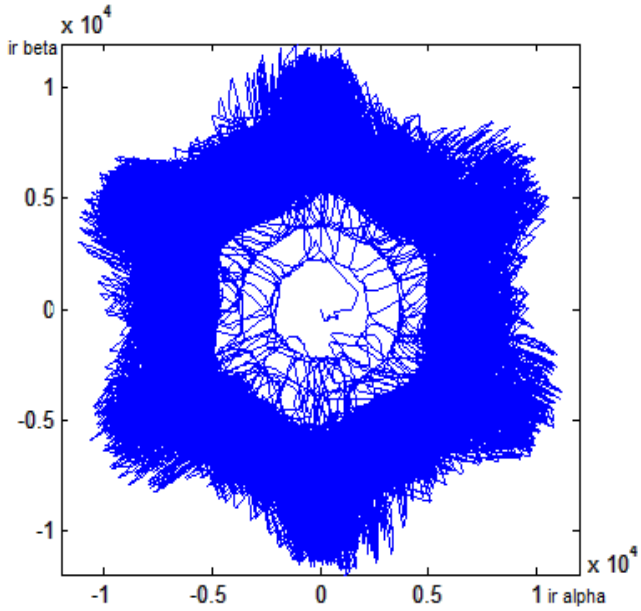


Figure 25: DFIG Rotor Currents Lissajous Curves in Case of Open Stator Supplying Phase 'c'.

The figures 24 and 25 represent in that order stator and rotor Lissajous curves. Curves are deformed comparing to reference Lissajous curves the case of defects-free operation. Also, the stator beta currents reach higher values than alpha currents as curves when phase 'a' is opened, but in this case the stator beta currents have higher values.

As showed in figure 25, rotor currents Lissajous curves have kept the same shape as case of open phase 'a' and phase 'b'.

We summarize that, the opening of one feeding DFIG stator phase generates an increase of the currents in the other phases and overheating of generator. Indeed, the resulting FFT spectrums analysing of stator currents help determine the amplitude of the harmonic components. These peaks amplitude, give us an idea about the defects severity affecting the wind energy chain.

Furthermore, the stator and rotor Lissajous curves are deformed according to defected generator supplying phase. So, we can detect DFIG defect types and know exactly which feeding phase is faulty.

6. Diagnosis Method Applied in Case of Defected Rotor Side Converter

Situated between the wind turbine generator and distribution networks, the back-to-back power converter has to satisfy both the generator side and grid side requirements [24]. As mentioned in paragraph 2.1, the control performed is indirect stator field oriented control. This way, DFIG is controlled by the rotor side converter. Owing to high voltage and a large transient current in the rotor windings, the power converter switching devices may be damaged due to their low power ratings of the switching devices [25]. So, connected to rotor circuit, a defected converter can cause generator damages. Hence, to diagnose DFIG rotor defects we apply a DFIG diagnosis method in case of open rotor side converter legs.

The figure below presents rotor side converter RSC components.

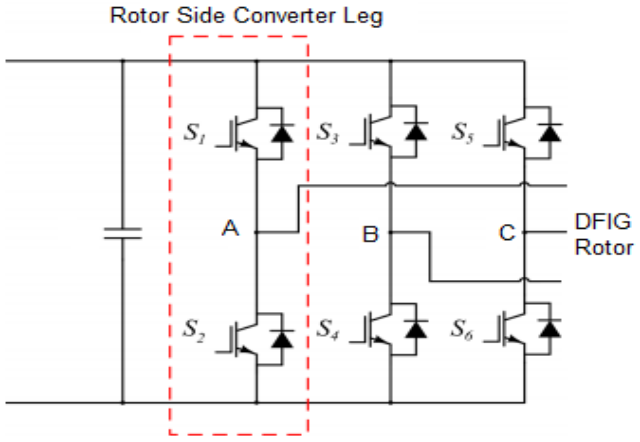


Figure 26: Rotor Side Converter with 3 Legs.

In the following part, we applied diagnosis method to DFIG in case of one rotor side converter leg opening.

The figures hereafter present the simulations results obtained in Matlab Simulink when RSC legs respectively (S1, S2), (S3, S4) and (S5, S6) are opened.

6.1. Open rotor feeding phase 'a'

The figure above exposes the FFT spectrum of stator currents in the case of opened RSC leg (S1, S2). The fundamental magnitude value reaches 4782, and THD is 89.45%. Comparing these values to reference frequency spectrum in the case of DFIG defects free operation, we note that the fundamental magnitude exceeded the reference fundamental magnitude. Also, we observe an absence of harmonics at 150 Hz and 250 Hz due to defected RSC.

In term of total harmonic distortion, his value passed from 0.58% to 89.45% in case of one open feeding rotor phase. This high THD level of harmonics in stator currents waveforms cause increasing of DFIG temperature which reduce generator life and subsequently lead to the loss of generator.

The following figures show Lissajous curves respectively generator

stator and rotor currents in case of open rotor feeding phase ‘a’.

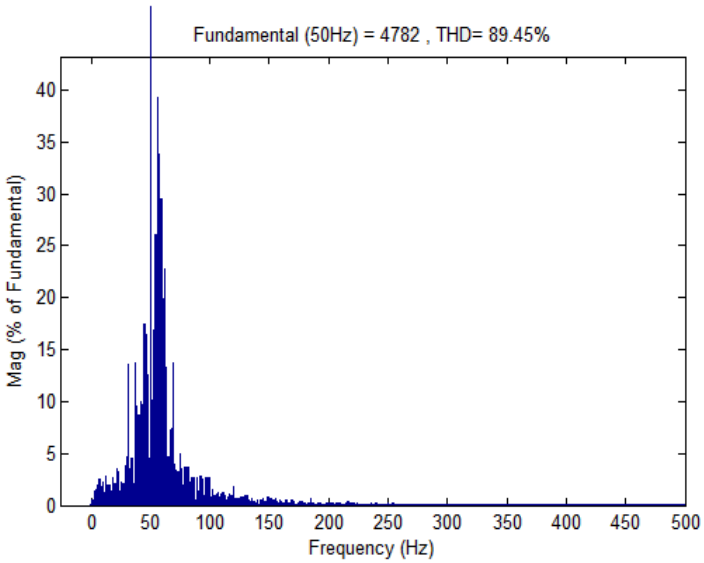


Figure 27: Stator Currents FFT Spectrum in Case of Open RSC Leg (S1, S2).

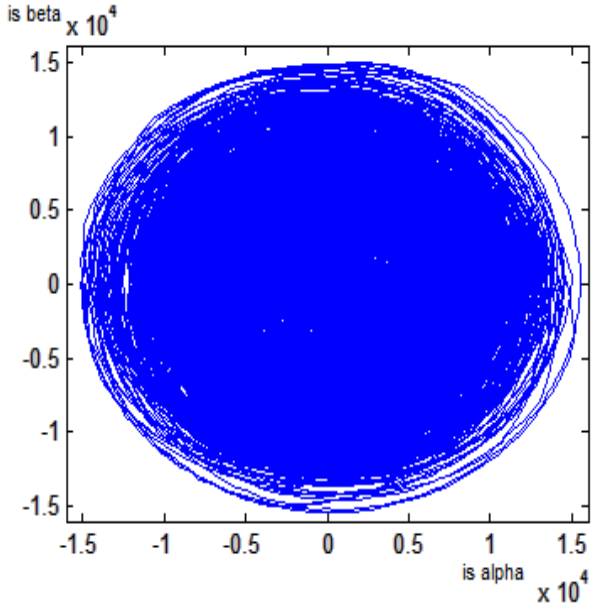


Figure 28: DFIG Stator Currents Lissajous Curves in Case of Open RSC Leg (S1, S2).

Comparing stator Lissajous curves obtained in cases of defects- free and defected generator, we see clearly in figure 28 that stator Lissajous currents curves have kept the circular shape whereas circumference and thickness increased because of stator currents increasing which makes harder the differentiation of three separate curves.

Lissajous rotor curves, as presented in figure 28, are totally deformed and the rotor currents increase. In fact, the opening of one rotor feeding phase (phase ‘a’) generated by opening of one rotor side converter (S1, S2) increase the currents in the other phases and cause the curves superimposition on each other. Which make impossible detection of the three separate curves as Lissajous curves obtained in case of generator under normal condition operating.

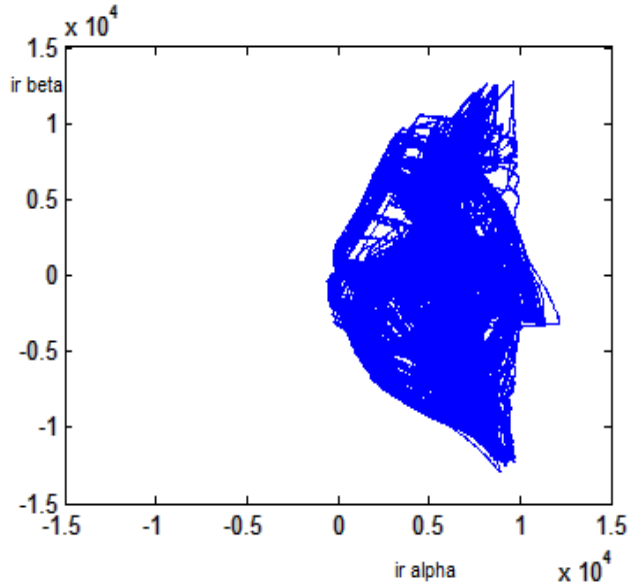


Figure 29: DFIG Rotor Currents Lissajous Curves in Case of Open RSC Leg (S1, S2).

Also, the curves have now two deformed petals flower shape instead of six petals obtained in reference data. The rotor currents ir alpha in this case are located on the positive part of Lissajous curves.

In the next part, we apply diagnosis method to DFIG in case of open RSC leg (S3, S4).

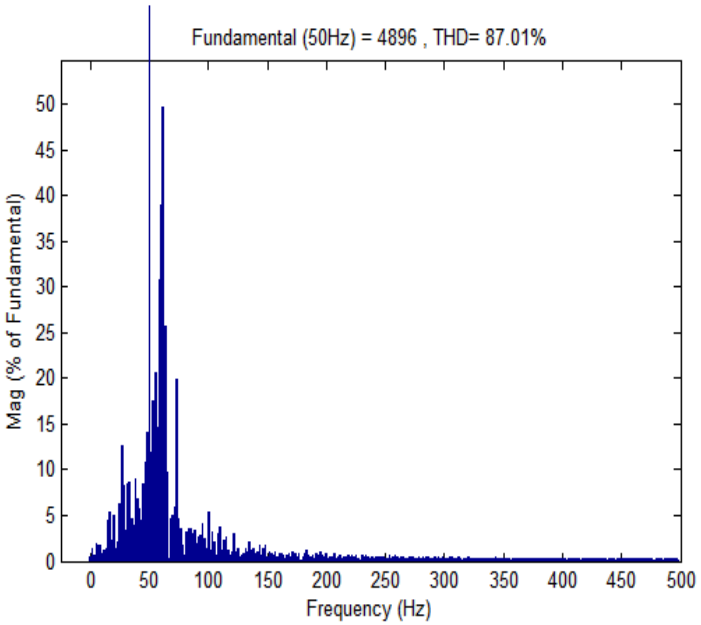


Figure 30: Stator Currents FFT Spectrum in Case of Open RSC Leg (S3, S4).

6.2. Open rotor feeding phase ‘b’

The figure 30 presents the FFT spectrum of stator currents in the case of open rotor feeding phase ‘b’. The fundamental magnitude reaches 4896 which is more than case of opened leg (S1, S2). The THD remains high with 87%. As the first case in paragraph 6.1 the harmonics at 150 Hz and 250 Hz did not appear in this frequency spectrum.

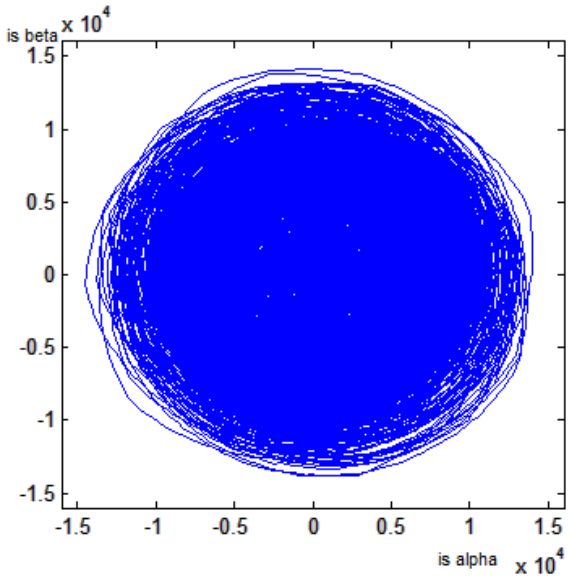


Figure 31: DFIG Stator Currents Lissajous Curves in Case of Open RSC Leg (S3, S4).

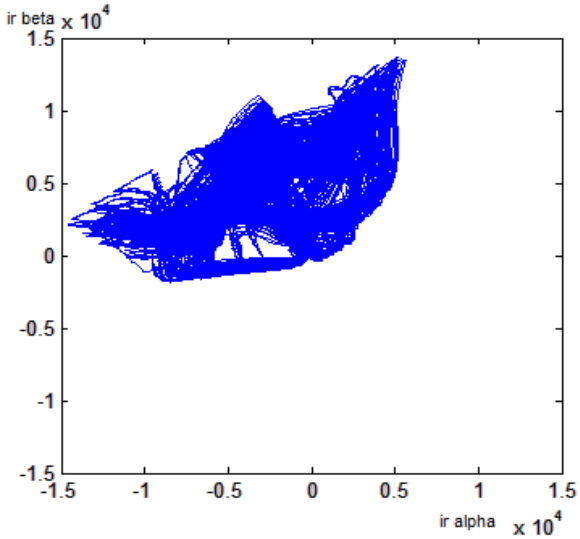


Figure 32: DFIG Rotor Currents Lissajous Curves in Case of Open RSC Leg (S3, S4).

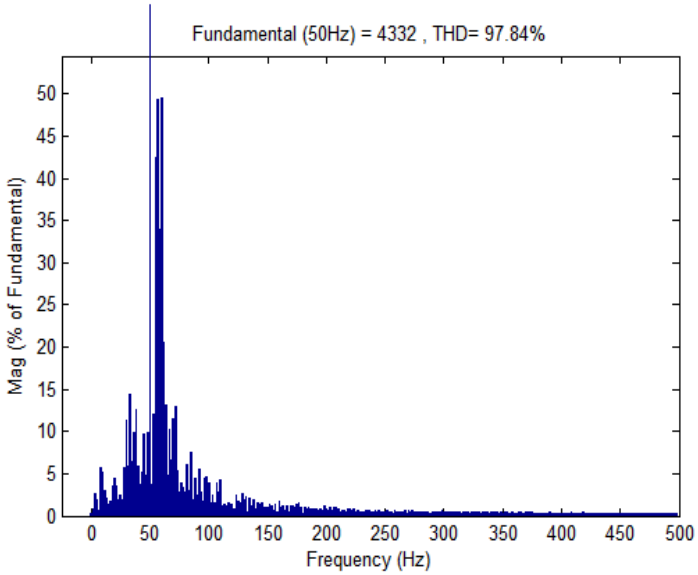


Figure 33: Stator Currents FFT Spectrum in Case of Open RSC Leg (S5, S6).

When rotor feeding phase 'b' is opened the stator Lissajous curves obtained in figure 31 have kept the circular shape. They are similar to the first defected rotor case.

For rotor currents, the opening RSC leg (S3, S4) as presented in figure 32, generates Lissajous curves with shape of two petals flower. But these two petals are inclined by $2\pi/3$ compared to the case of opening rotor phase 'a'. For the last part of this paper, our proposed DFIG diagnosis method is applied in case of open RSC leg (S5, S6).

6.3. Open rotor feeding phase 'c'

In the case of open rotor feeding phase 'c', we obtain the FFT spectrum of stator currents showed in figure 33. The fundamental magnitude reaches 4332 which is less than case of opened legs (S1, S2) and (S3, S4).
The THD reaches a maximum percentage with approximately 98% comparing to the previous opened phases.
According to the FFT spectrum graph and as foregoing cases, we note that the harmonics at 150 Hz and 250 Hz did not figure out.

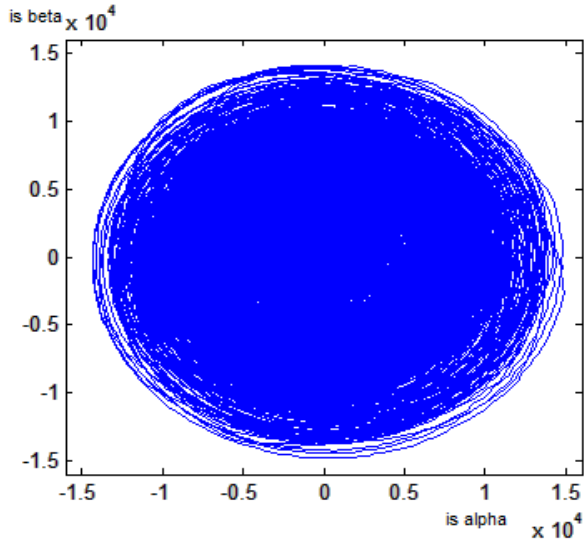


Figure 34: DFIG Stator Currents Lissajous Curves in Case of Open RSC Leg (S5, S6).

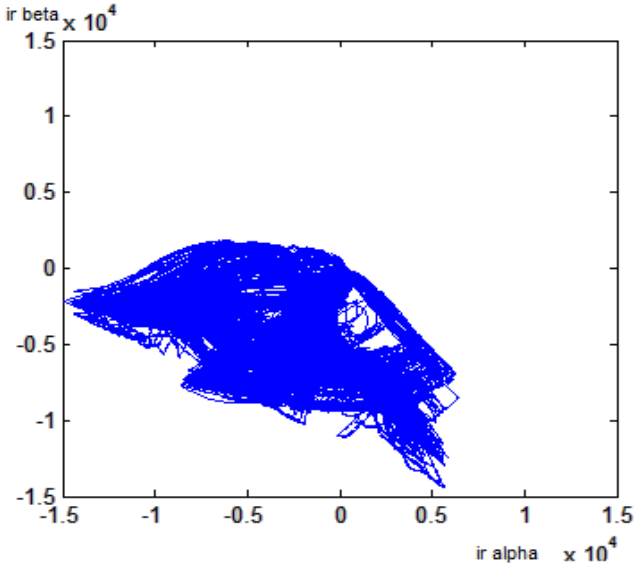


Figure 35: DFIG Rotor Currents Lissajous Curves in Case of Open IGBT Leg (S5, S6).

As opened feeding rotor phases 'a' and 'b', the stator Lissajous curves showed in figure 34 have similar shape. The generating Lissajous curves in figure 35 form two petals flower which are inclined by 4π/3 compared to the antecedent cases.

We deduct that, when one feeding DFIG rotor phase is opened the generator currents in the other phases increased. These high currents cause overheating of generator and can lead to the loss of generator. Through the frequency spectrums analysis of stator currents we determine the peaks amplitude, which show us the severity of the DFIG defects. And thanks to the stator and rotor Lissajous curves we detect exactly which feeding phase is faulty by means of deformation curves.

7. Conclusion

In this research paper, the method to diagnose wind turbine DFIG defects is developed through studying and analysing frequency spectrum and Lissajous curves of generator stator and rotor currents.

We start with creating wind energy conversion chain model in Matlab Simulink. Then, we apply indirect field oriented control to have separate and dynamical control on system flux and torque. Besides, the IFOC granted to have optimal, easy and efficient generator system operation over all the speed levels. Afterwards, based on studying generator currents frequency spectrum analysis and Lissajous curves, the method was developed with the aim of diagnosing and anticipating the potential DFIG wind turbine defects.

In the first step, this method was applied to generator under normal operating condition to get reliable reference data of generator behaviour.

In the second step, the technique was performed in case of open stator and rotor feeding phases. The results were compared with reference data obtained in case of defect-free operation.

Actually, according to defects types and by using Fast Fourier transform analysis, peaks appear or disappear at defined frequencies. Also, the maximum amplitude spectrum of the wavelet coefficients increasing, gives us an idea about the defects severity. Moreover, stator and rotor currents Lissajous curves are deformed considering shapes: thickness, circumference and direction of Lissajous curves in case of normal condition operating generator. We deduct that Lissajous curves are deformed according to the kind of defects attacking generator in wind turbine. Hence, the proposed DFIG diagnosis method allows recognition and prevention of potential defects attacking wind turbine.

Using Matlab Simulink, all the simulations had been realized. These results confirmed the efficiency and effectiveness of the proposed DFIG diagnosis defects method in the wind energy field.

References

[1] Global Wind Energy Council GWEC, "Global Wind Energy Report" April 2019.

[2] Moroccan Investment Development Agency, Kingdom of Morocco.

[3] R. M. R. Muthu, "Doubly Fed Induction Generator for Wind Energy Conversion System - a Survey", International Conference on Energy Efficient Technologies for Sustainability, pp. 617-628, 2013.

[4] M. A. El-Sharkawi, "Wind Energy an Introduction", University of Washington, Seattle, USA: 219, 2016.

[5] Outhred, Hugh & R. Bull, Stanley, "Meeting the Challenges of Integrating Renewable Energy into Competitive Electricity Industries", 2007.

[6] T.R. Ayodele, A.A. Jimoh, J.L Munda and J.T Agee, "Challenges of Grid Integration of Wind Power on Power System Grid Integrity: A Review", International Journal of Renewable Energy Research", Vol.2, No.4, 2012.

[7] K. Loudiyi, et al., "Grid Code Status for Wind Farms Interconnection in Northern Africa and Spain: Descriptions and Recommendations for Northern Africa", Renewable and Sustainable Energy Reviews, Elsevier, vol. 81, part 2, pp. 2584-2598, 2018.

[8] Hansen, et al., "Grid Fault and Design-Basis for Wind Turbines - Final Report", Roskilde, Risø National laboratoriet for Bæredygtig Energi, Denmark, Forsknings center Risoe. Risoe-R, no. 1714, 2010.

[9] M. Abomahdi and M.Srivastava, "Control of Grid Voltage and Power of Doubly Fed Induction Generator wind turbines during grid faults", IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE), vol. 9, iss. 4, pp 12-21, 2014.

[10] F. El Hammouchi, L. El Menzhi and A. Saad, "Wind Turbine Doubly-Fed Asynchronous Machine Diagnosis Defects-State of the Art", DEStech Transactions on Environment Energy and Earth Sciences, pp. 300-306, October 2017.

[11] F. El Hammouchi, L. El Menzhi, A. Saad, Y. Ihedrane and B. Bossoufi, "Wind Turbine Doubly-Fed Asynchronous Machine Diagnosis Defects -Part two", 2018 IEEE 5th International Congress on Information Science and Technology (CiSt), Marrakech, pp. 449-454, 2018. DOI: 10.1109/CIST.2018.8596498

[12] M. Allam, B. Dehiba, M. Abid , Y. Djeriri and R. Adjoudj, "Etude Comparative entre la Commande Vectorielle Directe et Indirecte de la Machine Asynchrone à Double Alimentation (MADA) Dédiée à une Application Eolienne", Journal of Advanced Research in Science and Technology (JARST),vol. 1, no. 2, pp. 88-100, 2014.

[13] Y. Ihedrane, C. El Bekkali, and B. Bossoufi, "Power Control of DFIG-Generators for Wind Turbines Variable-Speed", International Journal of Power Electronics and Drive System, vol. 8, no. 1, pp. 444-453, 2017.

[14] F. El Hammouchi, L. El Menzhi and A. Saad, "Wind Turbine Double-Fed Asynchronous Machine Diagnosis Defects-Part One", International Congress of Industrial Engineering and Systems Management CIGIMS, Meknes, Morocco, pp. 380-383, Mai 2017.

[15] A. Joshuva and V. Sugumaran, "Fault Diagnostic Methods For Wind Turbine: A Review", ARPN Journal of Engineering and Applied Sciences, vol. 11, no. 7, pp. 4654-4668, 2016.

[16] L. M. Popa, B. B. Jensen, E. Ritchie and I. Boldea, "Condition monitoring of wind generators", 38th IAS Annual Meeting on Conference Record of the Industry Applications Conference, 2003., Salt Lake City, UT, USA, vol. 3, pp. 1839-1846, 2003. Doi: 10.1109/IAS.2003.1257819

[17] E. ElBouchikhia, V. Choqueuse and M. Benbouzid,"Induction Machine Faults Detection Using Stator Current Parametric Spectral Estimation", Mechanical Systems and Signal Processing, vol. 52–53, pp. 447-464 , 2015.

[18] H. A. H. Al-Khazali and M. R. Askari, "Geometrical and Graphical Representations Analysis of Lissajous Figures in Rotor Dynamic System", Journal of Engineering (IOSR), vol. 2, issue. 5, pp. 971-978, 2012.

[19] L. El Menzhi, and A. Saad, "Lissajous Curve of an Auxiliary Winding Voltage Park Components for Doubly-Fed Induction Machine Electrical Faults Diagnosis", The Journal of Advanced Materials Research, vol. 860-863, pp. 2223-2231, 2014.

[20] F. El Hammouchi, L. El Menzhi, A. Saad, Y. Ihedrane, and B. Bossoufi, "Wind Turbine Doubly-Fed Asynchronous Machine Diagnosis Defects Using Stator and Rotor Currents Lissajous Curves", IJPEDS International Journal of Power Electronics and Drive System, Vol. 10, No. 2, pp. 961-970, 2019.

[21] A. Gupta , Dr. A. Shandilya, "Challenges of Integration of Wind Power on Power System Grid : a Review", International Journal of Emerging Technology and Advanced Engineering ,vol. 4, iss. 4, April 2014.

[22] A. Mouahid, A. Khamlich, A. Saad, "Study Of The Behaviour Of Wind Turbines Using Doubly Feed Asynchronous Generator During Voltage Dip", Mediterranean Telecommunication Journal, vol. 5, no. 2, pp. 126-132, june 2015.

[23] F. El Hammouchi, L. El Menzhi and A. Saad, "Diagnosis Method for Wind Turbine Doubly Fed Induction Generator under Grid Defects", International Journal of Information Science & Technology – IJIST, vol. 3, no. 3, pp. 46-55, 2019.

[24] F. Blaabjerg, M. Liserre, and K. Ma, " Power Electronics Converters for Wind Turbine Systems", IEEE Transactions on Industry Applications, vol. 48, no. 2, pp. 708-719, 2012

[25] J. J. Justo and F. A. Mwasilu, "Low Voltage Ride through Enhancement for Wind Turbines Equipped with DFIG under Symmetrical Grid Faults", Tanzania Journal of Engineering and Technology, vol. 37, no. 2, pp. 125-136, December 2018.

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 186-192 (2019)

www.astesj.com

Special Issue on Clean and Green Energy

ASTESJ

ISSN: 2415-6698

The Impact of Using Upgraded Biogas on Generator Performance

Wilson Babu Musinguzi*, Ibrahim Luqman Mpungu

Department of Mechanical and Manufacturing Engineering, Faculty of Engineering, Kyambogo University, P.O. Box 1 Kampala, Uganda

ARTICLE INFO

Article history:
Received: 27 April, 2019
Accepted: 13 September, 2019
Online: 08 October, 2019

Keywords:
Upgraded Biogas
Thermal Characterization
Efficiency

ABSTRACT

This research focussed on assessing the impact of using upgraded biogas from chicken waste on the performance of a gasoline generator. In the study, the characteristics of raw biogas were determined and a biogas upgrading device was developed before assessing the generator performance on upgraded biogas. A Geotech GA 2000 plus gas analyser was used for biogas characterization and a slightly modified Elemax SH 2900 gasoline generator (5 HP) was used. The findings revealed that biogas from chicken waste comprised of 57% CH_4 and 41% CO_2 with small amounts of H_2S, O_2 and NH_4. During biogas upgrading, the highest solvent to biogas flow rate used was 0.80. NaOH excelled with 93.0% CO_2 reduction followed by KOH at 82.5% and lastly water at 63.0%. Using upgraded biogas improved the generator's brake thermal efficiency from 14.2% for raw biogas to 17.6% for upgraded biogas. A locally fabricated scrubbing device significantly upgraded the biogas quality.

1. Introduction

Biogas is a renewable energy gaseous fuel produced from anaerobic digestion of biological wastes such as cattle dung, vegetable waste, sheep, poultry droppings, municipal solid waste, industrial waste water, land fill, etc. in the absence of oxygen and the presence of anaerobic microorganisms [1]. The process is achieved at elevated temperature range of 30 to 60°C. The produced biogas comprises of mainly methane (CH_4) and carbon dioxide (CO_2) with smaller amounts of hydrogen sulphide (H_2S) and ammonia (NH_4). Biogas is an excellent fuel for a large number of applications including production of heat and steam, electricity production, vehicle fuel and production of chemicals. In Uganda, the most common applications for biogas are cooking and lighting. Conventional gas burners and gas lamps can be modified to suit biogas by changing the air to gas ratio.

In this study, Okweru Poultry Farm was used as a case study. Apparently, biogas from chicken droppings was being used for lighting the chicken houses, heating the brooder and cooking by the farm workers. The researchers explored alternative uses for the biogas produced with focus on electricity generation using a gasoline generator.

Although biogas can be used as a fuel, it has an extremely low energy density on volume basis. This is because of the presence of CO_2 at higher quantities that reduce its calorific value, flame velocity and flammability range compared with other fuels [2]. The high CO_2 levels in biogas can be lowered through biogas upgrading. Global upgrading technologies include: absorption, membrane, absorption and cryogenic technology. Applying these techniques to small scale biogas plants presents complexity and high costs of investment. This leads to utilization of biogas without upgrading for basic applications in developing countries.

The utilization of biogas in stationary engines is a typical case where contaminants have to be removed for better performance. In dual fuel operation, increasing the ratio of biogas increases the brake specific fuel consumption while decreasing brake thermal efficiency and the brake engine power of a Compression Ignition (CI) engine [3]. It was reported in [4] that an improved thermal efficiency from 26.2% to 30.4%, when there is 21% reduction of CO_2 in biogas used in a Spark Ignition (SI) operation while a 15% CO_2 decrease in the biogas increased the thermal efficiency to 22% for dual fuel operations. However, there are other factors that influence the performance of a gas engine.

Various factors responsible for the low thermal efficiency of engines run on biogas were reported in [2]. These include; the long pilot ignition delay and high self-ignition temperature of biogas. These factors delay the fuel combustion process while the low burning velocity due to presence CO_2 leads to incomplete combustion. This increases the brake specific energy consumption and exhaust gas temperature of dual fuel modes.

*Corresponding Author: Wilson Babu Musinguzi, Faculty of Engineering Busitema University, P.O. Box 236 Tororo, Uganda, Tel: +256774076181, Email: Wilson.Musinguzi@gmail.com

This research aimed at ascertaining the performance improvement of a generator when upgraded biogas was used. This was carried out by analysing the characteristics of raw biogas and later developed a biogas upgrading unit. The performance of a generator was done using Model Elemax SH 2900 gasoline generator. Biogas upgrading reduces the levels of CO_2 which improves the burning velocity and calorific value of the gas among others. Through several experiments, the generator brake power output, brake specific fuel consumption and brake thermal efficiency were tested at different load profiles.

2. Materials and Methods

The experimental research included the determination of the biogas characteristics, biogas upgrading device design, fabrication and testing. Also experiments to assess generator performance were carried out for both raw and upgraded biogas. This yielded the primary data utilized in the analysis.

2.1. Characteristics of Raw Biogas

In line with biogas characterisation, a Geotech GA 2000 plus Gas Analyser was used to measure the percentage composition of the biogas constituents. The biogas pressure was measured in kPa using a pressure gauge while a thermocouple was employed for temperature measurement in degrees Celsius. The gas flow rate was measured in litres/minute using an Omega HHF11A Air flow meter.

2.2 Designing the Biogas Upgrading Device

The factors that influence the upgrading process include the dimension of the scrubbing tower, gas pressure, raw biogas composition, water flow rate and purity of the water in use. According to [5], it is possible to produce high quality CH_4 enriched gas from biogas by chemical absorption, where a packed bed column and bubble column are used to provide liquid to gas contact.

Therefore, a low-cost upgrading system was designed to receive raw biogas from the bottom and force it through packed marbles to reduce its CO_2 content. According to [6], for purification, the height to diameter ratio should be in the range 10:1 to 130:1. For experimental purposes, a device of height to diameter ratio of 10:1 was made and utilized in this study. Based on recommendations by [7], the column was divided into three parts. The top part is half the height of the bottom part while the middle part contains the packing. Marbles were used as packaging material and they occupied 70% of the column height. Based on the data above, the upgrading device was designed with dimensions as shown in Figure 1 while Figure 2 shows the assembly of the device.

2.3 Generator Performance

In internal combustion engines, the combustion of a fuel occurs with an oxidizer (usually air) in a combustion chamber that is an integral part of the working fluid flow circuit [2]. Petrol engines can run fully on biogas whereas diesel engines require combination of biogas and diesel. In the diesel engine, the primary gaseous fuel is inducted with air into the engine cylinder and a small amount of diesel, called pilot fuel is injected for promoting combustion [8].

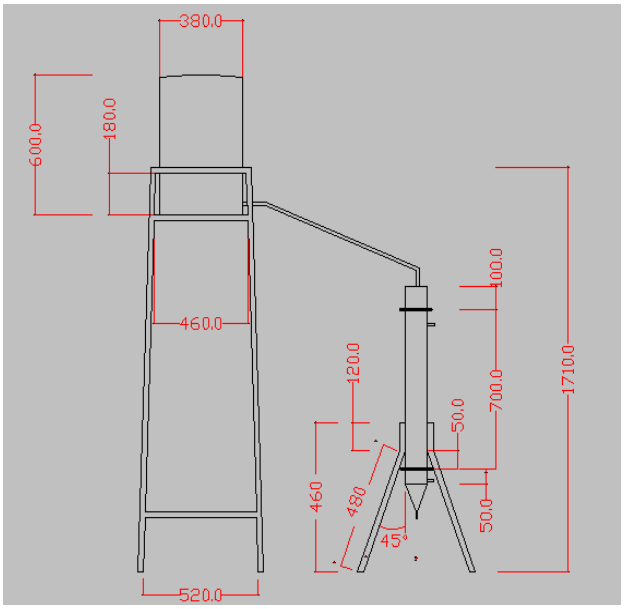


Figure 1: Design of the upgrading device (units in mm)

Figure 2: Assembly of the biogas purification device

To assess the generator performance, an Elemax SH 2900 gasoline generator was used. Elemax SH 2900 generator was used because it has a spark ignition engine which can easily ignite the biogas using the spark plug unlike compression ignition engines. Therefore, biogas could be used as a fuel without blending it as it is with compression ignition engines. Table 1 gives the generator specifications.

The gasoline generator was slightly modified to enable it run on biogas. This was done by inserting a wooden piece holding a 10mm metal duct in between the generator engine and carburettor. This duct was to direct biogas to the generator engine without necessarily changing its entire make-up. Therefore, as the generator run, the fuel could be swiftly changed from gasoline to biogas.

Table 1: Elemax SH 2900 Generator Specifications

S/N	Parameter	Specification
1	Model	Honda GX160 SH 2900 Generator
2	Engine type	4-stroke OHV
3	Rated power	3.6 kW @ 3600rpm
4	Cooling system	Air cooled
5	Ignition system	Transistorized magnets
6	Generator AC output	2400VA @ 220V, 50Hz

During experiments, biogas was collected in a polythene bag of capacity 1m^3. The generator was manually started on petrol and later biogas introduced. The supply of petrol was deliberately and steadily reduced as biogas was steadily introduced while regulating the air intake. Eventually, petrol was completely cut off and the generator ran on biogas. The experimental setup is shown in Figure 3.

The experiments were conducted under different loading conditions. Compact fluorescent lamps were used as the test loads. These test loads were grouped in percentage of the total load which was made of 16 lamps connected in series and parallel. Test load one comprised of 25% of the total load, load two was made of 50% while loads three and four were made of 75% and 100% of the total load respectively.

For each test load, various measurements were taken after stabilization. These include; current, voltage and fuel consumption. An ammeter connected in series was used for current measurement while a voltmeter connected in parallel with the test load was used for voltage measurement. Current was measured in amps while voltage was measured in volts. An air flow meter was used to measure the biogas flow for a given period of time. The results were used to determine the biogas fuel consumed by the generator for each test load given the period it was run. A ball valve connected to the pipe on the biogas bag was used to control the flow rate. The results obtained for each test load were used in calculation of Brake Power, Brake Specific Fuel Consumption (BSFC) and Brake Thermal Efficiency (BTE), using the formulae obtained from [2].

Engine brake power output (Pb):

$$\text{Voltage developed (V) x Current produced (I)} \quad (1)$$

Brake specific fuel consumption (BSFC):

$$\frac{\text{Fuel Consumption (FC)}}{\text{Engine brake power output (pb)}} \quad (2)$$

Brake thermal efficiency (BTE):

$$\frac{3600 \text{ x pb}}{\text{FC x LHV}} x\,100\% \quad (3)$$

Where LHV: Lower Heating Value of the gas.

Table 2 below shows the thermal content of biogas that was considered when calculating the Brake Thermal Efficiency. Since CH_4 is the only combustible gas in biogas, it was used when considering the lower heating value of the gas. From the biogas characterization done and the results obtained from the upgrading process, the heating values of biogas were obtained from Table 2 below based on their CH_4 content.

Table 2: Methane Yield from Animal Waste [9]

Animal	Typical experimental yield/kg manure	CH_4%	CO_2%	Thermal content MJ/m^3
Cattle	200 – 350 L	57.5%	46.5%	23
Poultry	550 – 650 L	70.0%	30.0%	28
Pig	400 – 500 L	65.0%	35.0%	26

Figure 3: Experimental Setup for the Generator Running *on Biogas*

3. Results and Discussions

3.1 Characterization of Biogas from the Digester

Biogas from chicken droppings was got from a 30m^3 Fixed Dome Digester. It comprised of 57.3% CH_4, 41.6% CO_2, 0.2% O_2, 332 ppm of H_2S and a balance of 0.4%. These were read at a digester temperature of 28.1^{o}C, pressure of 16kPa and maximum flow rate of 15litres/minute. Biogas from poultry wastes can comprise 70% CH_4 (Abdul Kareem, 2005). However, the variations exhibited can be attributed to numerous factors including the operating temperature of the digester, level of substrate mixing, the pH, carbon to nitrogen ratio (C:N) and toxins in the feedstock which may be as a result of the vaccines administered to the chicken. Biogas production can be increased by operating the digester at thermophilic temperature ranges. The said digester was run at mesophilic temperature ranges expected to be between 25-35^{o}C. Factors such as, optimum substrate mixing, neutral pH and optimum C: N ratios can affect the biogas yield quantities and component distribution.

3.2 Experimental Results obtained from the designed Biogas Upgrading Device

To establish the performance of the fabricated apparatus, different experiments were carried out to determine the device efficiency. This was done using three solvents including; H_2O, NaOH and KOH. At solvent to biogas flow rates of 0.20, 0.40, 0.60 and 0.80, the data captured was relating to biogas composition, CH_4 enrichment and CO_2 reduction. The comparative data obtained for analysis of the three solvents using average values is given in Table 3. The findings in table 3 indicate that at the same ratio of solvent to biogas flow rate, the CO_2 reduction was lowest for water, followed by KOH and highest for NaOH. On the other hand, the same ratios of the flow rate for the solvent to biogas, highest CH_4 enrichment was generated in NaOH, followed by KOH and lowest for H_2O. This is because solutions of NaOH and KOH have enhanced scrubbing capabilities for CO_2 removal. Figure 4 illustrates the percentage CH_4 enrichment using the various solvents at varying flow rate ratios.

Graphically, for the three solvents, best performance was exhibited at a ratio of NaOH to biogas flow rate of 0.80 where CO_2 reduction clocked 93% and the CH_4 enrichment level was 60.6%. The device thus performed better at higher flow rate ratios of 0.80 than at low flow rate ratios of 0.20 for all solvents. This is because there is more CO_2 absorption when the volume of solvent interacting with a given biogas volume is relatively high. Although better results are obtained at high flow rates, further increasing flow rate led to flooding of the device. Higher columns can properly cater for higher flow rates hence delaying flooding. The height of the column can be technically increased by packing. Packaging materials greatly increase the surface area of the column and hence interaction time of the gas and solvent thereby delaying flooding. Packaging further allows for efficient contact between the water (solvent) and gas phases in a counter current absorption process [7]. It is recommended that almost 70% of the column height should be filled with packaging material for good results [10]. However, the kind of packaging material defines the impact on the performance of the column. This is because packaging material affects the uniform distribution of the solvent and biogas as they flow across the column cross section.

3.3 Characteristic Generator Performance

Having developed and tested the performance of the biogas upgrading unit, control experiments were conducted to establish the performance of biogas in electricity generation. This was done by replacing the generator fuel from gasoline with biogas. Both raw and upgraded biogas was used and varying results were obtained.

3.3.1. Experimental Results of the Generator Performance on Raw Biogas

In this experiment, raw biogas was used in the generator. Ammeter, voltmeter and flow meter readings were recorded and used in Formulae 1 to 3, used to come up with data in Table 4. From the biogas characterization done, the CH_4 content was recorded at 57.3%. Therefore, the heating value of the raw gas was assumed to be 23MJ/m^3 or 20MJ/kg as earlier noted in Table 2.

3.3.2 Experimental Results of the Generator Performance on Upgraded Biogas

In this set of experiments, biogas was upgraded before it was used in the generator. After the biogas upgrading process, the CH_4 content was enriched to 78% using plain water as the solvent at water to gas flow rate ratio of 0.80. Therefore, the heating value of the upgraded biogas was assumed to be 28MJ/m^3 or 24.3MJ/kg since it had a higher CH_4 content than that quoted for poultry wastes with 70% CH_4 in Table 2. Ammeter, voltmeter and flow meter readings were recorded and used in Formulae1 to 3, and used to come up with data in Table 5. From Figure 5, the highest brake power is achieved at 100% loading of the generator with upgraded biogas recorded at 0.239 kW while the lowest brake power was attained at 25% electric load with raw biogas recoded at 0.060 kW. The current consumed increases with increase in the connected load. Therefore, the more loads connected, the higher the current consumed and hence an increase in brake power. This is in line with [2] where it is stated that increase in loading increases the combustion quality of the fuel and hence the power output of the generator. This is because the generator output depends on its fuel burning efficiency. From Figure 6, Brake Specific Fuel Consumption (BSFC) decreased with percentage electric load. This was because BSFC decreases with increase in electric load as it depends on fuel consumption directly and brake power inversely which increases with increase in brake load [2]. BSFC is higher for raw biogas at 2.07 kg/kWh as compared to that of upgraded biogas at 1.45 kg/kWh. This is because a larger volume of biogas was required for the same load for raw biogas than for upgraded biogas.

Upgraded biogas had a higher calorific value of 24.3 MJ/kg due to the high CH_4 content of 78% and hence the less volume of biogas required for the same load. It is well known that the calorific value of the fuel has an impact on the amount of fuel required for running the engine. On the other hand, Brake Thermal Efficiency (BTE) of upgraded biogas is higher than that from raw biogas. This is because a drop of CO_2 in biogas for dual fuelling increases the thermal efficiency. The highest value of upgraded biogas BTE was 17.6% compared to 14.2% of raw biogas and the lowest value for upgraded biogas was 10.2% as compared to 8.7% of raw biogas. BTE increases with increase in brake power. It can be noticed from Figure 6 that BTE increased with increased load. This is the same pattern demonstrated by brake power in Figure 5 since it's directly proportional with the percentage load. BTE further depends on brake power, fuel consumption rate and calorific value of biogas. Since upgraded biogas had a higher calorific value of 24.3MJ/kg compared to raw biogas' 20 MJ/kg and lower fuel consumption, it greatly contributed to the higher efficiency of upgraded biogas.

Table 3: Comparison of the Three Solvents on CH_4 Enrichment and CO_2 Reduction

Flow Rate Ratio	H_2O		KOH		NaOH	
	CH_4 enrichment	CO_2 reduction	CH_4 enrichment	CO_2 reduction	CH_4 enrichment	CO_2 reduction
0.2	18.3	42.8	26.4	54.8	43.5	76.4
0.4	23.2	48.6	34.0	65.4	51.1	84.9
0.6	30.7	58.7	40.1	72.8	57.6	91.8
0.8	36.8	63.0	48.9	82.5	60.6	93.0

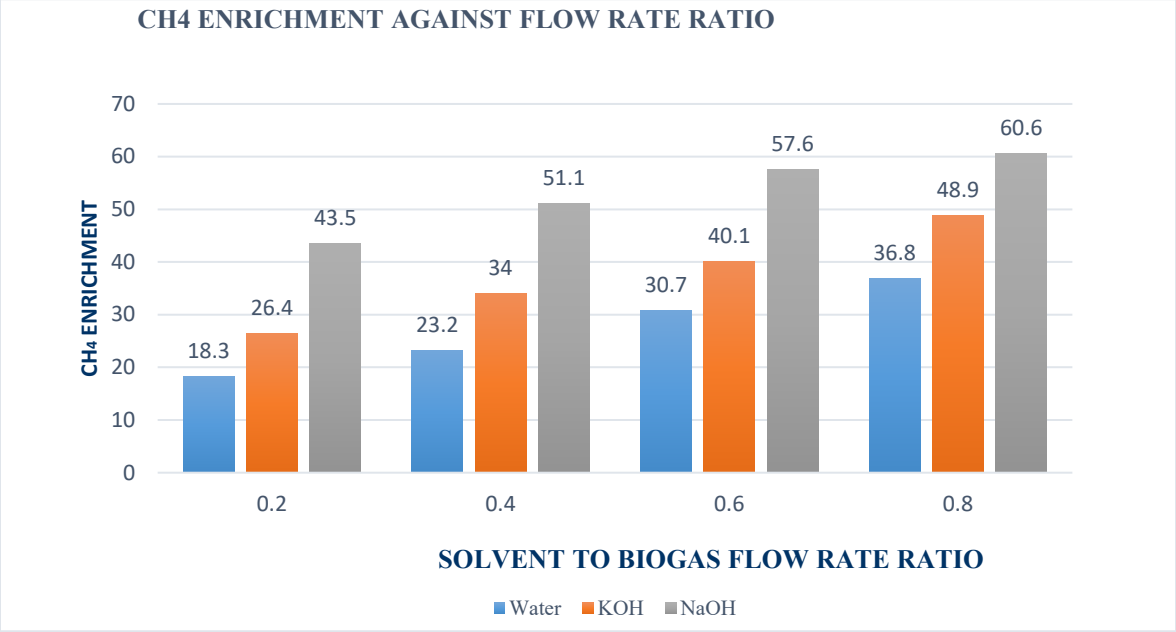


Figure 4: Comparison of the Three Solvents on CH_4 Enrichment

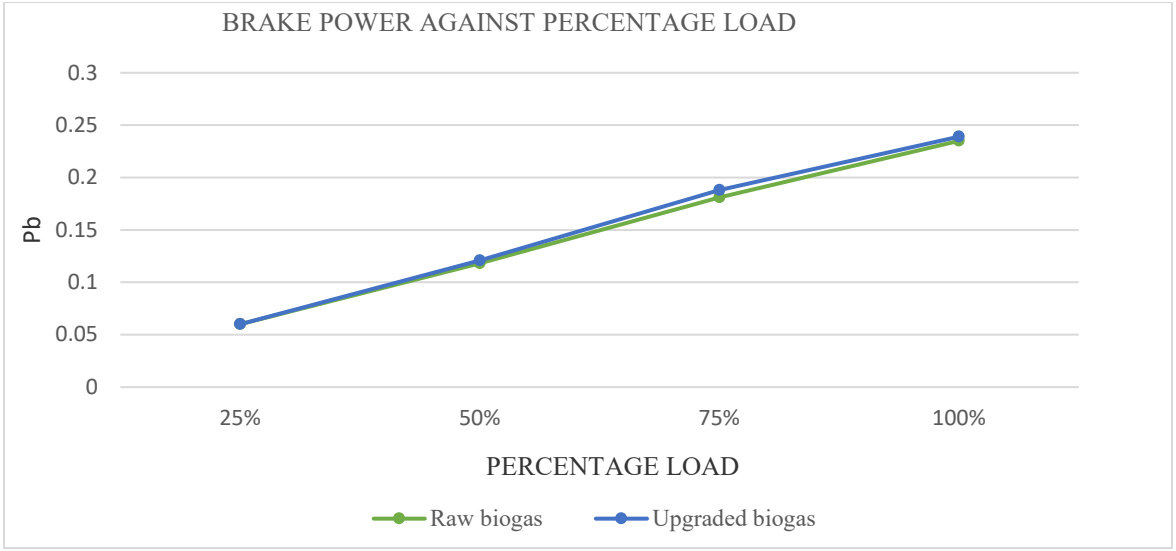


Figure 5: Variation of Brake Power against Percentage Load for both Raw and Upgraded Biogas

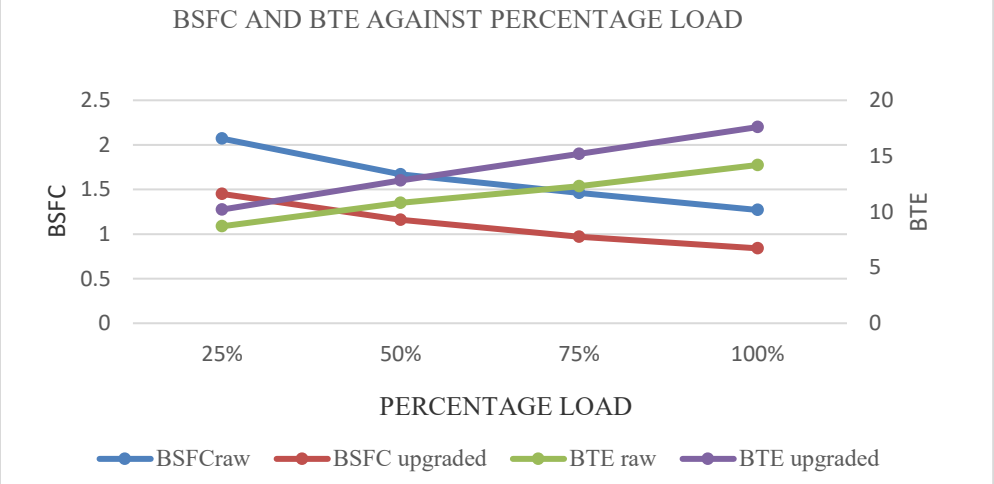


Figure 6: Variation of Brake Specific Fuel Consumption and Thermal Efficiency with Percentage Load

Table 4: Generator Performance on Raw Biogas

S/N	ELECTRIC LOAD	25%	50%	75%	100%
1.	Voltage (V)	222.2	227	226.3	223.8
2.	Current (A)	0.27	0.52	0.80	1.05
3.	Pb (kW)	0.060	0.118	0.181	0.235
4.	Fuel consumption (kg/h)	124.1	197.3	264.9	297.9
5.	BSFC (kg/kWh)	2.07	1.67	1.46	1.27
6.	BTE (%)	8.7	10.8	12.3	14.2

Table 5: Generator Performance on Upgraded Biogas

S/N	ELECTRIC LOAD	25%	50%	75%	100%
1.	Voltage (V)	225.9	232.7	235	227.6
2.	Current (A)	0.27	0.52	0.80	1.05
3.	Pb (kW)	0.061	0.121	0.188	0.239
4.	Fuel consumption (kg/h)	88.6	140.0	183.2	201.2
5.	BSFC (kg/kWh)	1.45	1.16	0.97	0.84
6.	BTE (%)	10.2	12.8	15.2	17.6

During the start of experiments, the generator was run on petrol to heat up and later biogas introduced. This was because it was hard burning biogas at the generator start due to the high heating temperatures required. However, this wasn't the case when the generator had been running because the generator would have heated up already. This further highlighted the impact of biogas' low heating value. BTE increases with increased load due to increase in combustion zone temperature. Increase in load increases the combustion zone pressure and heat release rate. This improves the fuel combustion quality hence increasing the power output of the generator. Therefore, the higher the generator power output, the better the BTE.

4. Conclusion

The performance of the biogas upgrading device not only depends on the biogas flow rate but also depends on the dimensions of the scrubbing tower, biogas pressure, packaging material and purity of the water (solvent) used in the process. Upgraded biogas had a higher calorific value of 24.3 MJ/kg due to the high CH_4 content of 78% and hence the less volume of biogas required for the same load. Increase in load increases the combustion zone pressure and heat release rate. This improves the fuel combustion quality hence increasing the power output of the generator. Therefore, the higher the generator power output, the better the BTE. Using upgraded biogas improved the generator's brake thermal efficiency from 14.2% for raw biogas to 17.6% for upgraded biogas.

Conflict of Interest

The authors declare that although they may have conferred with others in carrying out the research and preparation of this paper, and drawn upon the range of sources cited in this work, this paper is their original piece of work and has not been published in any other journal.

Acknowledgment

The authors gratefully acknowledge the funding and support received from RELOAD, a Germany organization dedicated to finding solutions to post-harvest losses, and Dr. Peter Okweru of Okweru Poultry Farmers for availing his farm to act as the study area where all experiments were carried out.

References

[1] S.S. Kapdi, V.K. Vijay, S.K. Rajesh, R. Prasad, "Biogas scrubbing, compression and storage: perspective and prospectus in Indian Context" Ren. Energ., 30(8), 1195–1202, 2005. http://dx.doi.org/10.1016/j.renene.2004.09.012

[2] A.V. Reddy, T. Sharath-Kumar, D,K, Tharun-Kumar, B. Dinesh, Y.V.S. Sai-Santosh, "Energy and exergy analysis of I.C. engines" Eng. and Sci., 3(5), 7– 26, 2014. ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805

[3] N.H.S. Ray, P.R. Swain, M.K. Mohanty, "An investigation on performance characteristics of C.I. engine using biogas and diesel in dual fuel mode" Sci., Eng. and Technol. Research, 3(6), 1716 – 1723, 2014. ISSN: 2278 – 7798

[4] D. Barik, S. Sah, S. Murugan, "Biogas production and storage for fueling internal combustion engines" Emerging Technol. and Advanced Eng., 3(3), 193–202, 2013. ISSN 2250 – 2459

[5] L.B. Allegue, J. Hinge, "Biogas and bio-syngas Upgrading," Report, Danish Technological Institute, Kongsvang Allé 29 Dk–8000 Aarhus C., 2012.

[6] A.K. Dubey, "Wet scrubbing of carbon dioxide," Annual Report of Ciae, Bhopal, India, 2000.

[7] R. Chandra, V.K. Vijay, P.M.V. Subbarao, "Vehicular quality bio-methane production from biogas by using an automated water scrubbing system," ISRN Ren. Energ., 1-6, 2012.

[8] N.H.S. Ray, M.K. Mohanty, R.C. Mohanty, "A study on application of biogas as fuel in compression ignition engines," Innovations in Eng. And Technol., 3(1), 239 – 245, 2013. ISSN: 2319 – 1058

[9] A.S. Abdul-Kareem, "Refining biogas produced from biomass: an alternative to cooking gas," Leonardo Sci., 7, 1–8, 2005. ISSN 1583-0233

[10] D. Barik, S. Sah, S. Murugan, "Biogas production and storage for fueling internal combustion engines" Emerging Technol. and Advanced Eng., 3(3), 193–202, 2013. ISSN 2250 – 2459

[11] N. Tippayawong, P. Thanompongchart, "Biogas quality upgrade by simultaneous removal of CO_2 and H_2S in a packed column reactor," Energ., 35(12), 4531–4535, 2010. DOI: 10.1016/j.energy.2010.04.014.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 193-206 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Prediction of Non-Communicable Diseases Using Class Comparison Data Mining

Ammar Al-Dallal[*,1], Amina Al-Moosa[2]

[1]*Ahlia University, Computer Engineering Department, P.O. box 10878, Bahrain*

[2] *Ahlia University, Information Technology Department, P.O. box 10878, Bahrain*

ARTICLE INFO

Article history:
Received: 30 April, 2019
Accepted: 07 September, 2019
Online: 08 October, 2019

Keywords:
Communicable disease
Data mining
Diabetes
Predictive data mining
Heart disease
Class comparison

ABSTRACT

Data mining is recognized as an effective technique for extracting and retrieving valuable information or decision from the vast available data. Because of the nature of the functionality of medical centers and hospitals, their data centers contain a collection of valuable information about their patients. By properly processing these data, different applications can be developed to utilize them. These applications could participate in predicting and diagnosing particular diseases. Two prime diseases realized to impact the overall health of society are heart diseases and diabetes. The presented work intends to develop and test a software application that helps doctors and practitioners predict the emergence of noncommunicable diseases (NCDs) such as diabetes and heart diseases. The application applies the predictive data mining model to the medical records which are collected from the Bahrain Defense Force Hospital (BDFH). The BDFH doctors evaluated the application and executed it on actual patients. The results obtained are accurately matching the expectation of doctors in BDFH. All kinds of risks are categorized appropriately according to the defined categories. As a conclusion, this application can help doctors in making proper decisions toward patient health risks. In addition, data mining is more supportive for the health sector and is essential for exploring the knowledge to be used in the health care sector.

1. Introduction

The term data mining refers to the process of knowledge discovery discipline in databases (KDD). It is associated with various computing fields; such as databases, artificial intelligence, as well as software engineering. It can also be used to overcome various problems that emerge where a large volume of data is involved.

The World Health Organization (WHO) is the world body that monitors international health. It showed that 68% of all mortality around the globe is due to noncommunicable diseases (NCDs), and the major killer NCDs are cardiovascular diseases (CVDs), diabetes, lung diseases, and cancer [1-2]. It added that heart diseases cause 12 million deaths globally. CVD comprises of various heart-related disease and its functioning problems that contribute significantly to the adult mortality in some countries.

The (popular) disease or new-generation disease is diabetes. It has emerged as a consequence of inactivity, fewer movements, sedentary lifestyle, and unhealthy eating habits. Given the low activities and other lifestyle reasons, the pancreas produces insufficient insulin to control blood sugar, resulting in diabetes. In a small country in the population such as Bahrain, these two NCDs present a multifaceted problem that must be addressed [3]. Hospitals and health centers accumulate a massive amount of patient's data. This data can be entered or fed to a data mining engine to help in predicting and diagnosing various types of diseases.

The accurate and efficient prediction is a consequence of medical diagnosis. Therefore, it is an important task at that stage. Unfortunately, not all doctors hold expertise for various specialization. In addition, the number of specialized practitioners is also found to be insufficient in many health centers, especially when it is related to heart diseases. Hence, there is a great need to develop a system that assists in predicting and forecasting the patient's diagnosis. It would certainly bring plenty of relief and help doctors.

[1] Corresponding author: Ammar Al-Dallal, Computer Engineering Department, Ahlia University, Manama, Bahrain. Email: aaldallal@ahlia.edu.bh, as_aldallal@yahoo.com

This research is designed to assess the data of the patient concerning his health which is necessary to forecast NCDs using inferential data mining and KDD to assists the healthcare professionals in evaluating, as well as predicting, the recurrence or occurrence of NCDs. As a result, this research offers a platform for doing such a system to be able to develop all sets of rules and make them available to medical doctors and other practitioners to anticipate with a degree of correctness the prospect of NCDs among patients. To be able to detect the prime elements which are vital for NCD prediction, the development of an application takes place to compile the massive volumes of patient data for forecasting the potential future trends by employing data mining. Afterward, its testing occurs, which allows the physicians and experts to use the information pertaining to NCDs for improving its diagnosis and reducing the diseases. This project utilizes huge amounts of data obtainable from the BDFH to anticipate NCDs by employing data mining technique.

The contribution that the study adds is that it develops a software which predicts the diagnosis and is tested by BDFH actual practitioners. Using the software, the practitioners were gratified with the study results.

1.1. Purpose of the Study

The purpose of the study is to apply KDD and execute the inferential data mining technique to examine health data, which is essential for predicting NCD. This assists doctors in analyzing or even predicting the recurrence or occurrence of NCDs in patients. Currently, BDFH has no tools or techniques to undertake this task. Hence, a Predicting NCD Application (PNCDA) software will be developed by applying the inferential data mining technique on the compiled huge volumes of available vital data. This application will be available for doctors to be able to predict future trends with accuracy, for the NCD potential among the BDFH patients. Following it, the application will be assessed for demonstrating the model which enables the clinicians and other stakeholders to have access to this facility and use it during their diagnosis process to improve patients' health and to reduce such diseases.

Because of the sensitivity of this project as it deals with confidential patient data, two types of data will be utilized: secondary and primary data. Secondary data will be extracted from the records of the BDFH, while primary data are gathered in person where the related data are organized as per the software requirements. The approach of the interview will be used for collecting primary data, which will recruit about 30 percent to 40 percent of doctors, nurses, as well as BDFH paramedics.

1.2. Study Significance

The objective of this project is to experiment as well as assess different algorithms of data mining, which will assist in NCDs prediction, comparison, and contrast of the effective ways of predicting the disease. The experiment is assumed to serve as an instrumental tool for the doctors for predicting in complex medical cases related to NCDs and advice their parents as per the predictions generated through accurate algorithms, that will have a huge advantage for the field of medical science. This research would provide evidence that the technique of data mining is effective for physicians for predicting and forecasting risky medical cases as practiced across developed nations. This research is set to act as a predecessor for accumulating data for patients in the Kingdom of Bahrain in the future.

1.3 Study Limitations

The study has certain limitations such as it includes patients recorded in BDFH only in Kingdom of Bahrain. Once completed, the findings of the research would be communicated and discussed with the healthcare authorities in the kingdom, where additional comprehensive research can be conducted for covering the whole Bahrain population.

This whole research is categorized into five sections; where section two presents a comprehensive review of the data mining methods and its implementation in the medical discipline. Following it, section 3 provides an explanation of the process for data mining, which is used for the prediction system for NCD. It also provides a description of the prediction software, which is developed. The developed software for prediction and its results are reflected and analyzed in section four. Lastly, section five briefly summarize the overall findings of the research and provides a direction to the future researches for expanding the research horizon.

2. Theoretical Background and Literature Review

In the present times, an increased acceptance of the data mining on the international forums is being recognized, across different medicine and life spheres. Considering the dynamics scope of data mining related to its efficacy for enhancing the healthcare outcomes, this section intends to highlight the theoretical basis which assists in NCD determination as well as the application of the data mining in forecasting.

2.1. Non-communicable Disease

At present, the unbridled growth of NCDs is recognized as the primary cause of mortality across the world. It is reasoned that the increase in sedentary lifestyle and lack of exercise has added to its increase. As it is well-recognized that an active lifestyle is vital for proper maintenance and functioning of the human body. According to 2016 WHO statistics, more than 15 million deaths occurred as a result of NCD such as diabetes and cardiovascular disease [1]. The most surprising and alarming finding of this statistics was that almost half of the deaths were of individuals who were less than 70 years of age.

A. Diabetes

The statistical evaluation of World Health Organization (WHO) findings, and its comparison with the outcomes of 2014, it is found that diabetes had an increasing percentage of about 409 percent for 34 years, ranging from 1980 to 2014. The main contributor to the increasing percentage is the changing world dynamics where the living status of the majority of the countries has changed for both developing and underdeveloped countries [1]

As per the healthcare professionals, the disease of diabetes occurs when the gland of the body (pancreases) is able to release an adequate amount of insulin. It affects the human body function as insulin is responsible for carrying sugar from the bloodstream to numerous cells in order to be used as energy. The deficiency of insulin in the body makes its natural functioning difficult. As a

consequence, the body releases high levels of glucose in the urine. Its prevalence for a long-term can cause an organ to fail, cause CVD, and affect the other functions of the body. The deteriorative effects which emerge as a consequence of diabetes are the reason it is ranked as the fourth primary NCD by WHO [2]. It can also lead to the complication of CVDs. The WHO organization [2] documented that diabetes and CVDs were the reasons for more than 11 million fatalities that occurred before the age of 70, as shown in Figure 1.

The primary factor which contributes to the occurrence of diabetes includes the uncontrolled level of blood glucose. Using data mining techniques, the healthcare professionals across the world would be able to forecast illness in an effective manner and will be able to integrate better management technique for patients at high risk. This adds to the significance and needs for the disease analysis and predictions for overcoming it and providing relief to the majority of the patients.

Similar to other regions, Bahrain also lists diabetes as a major health concern. The prevalence of diabetes is found across different ages and populations comprising of different characteristics. The report by WHO for Bahrain documented that diabetes account to 13 percent of the deaths in the region. It also indicates the increasing detrimental outcomes of diabetes, which continue to grow at an increasing speed [3].

Generally, diabetes is categorized into two types; namely, type 1 and type 2. In type 1, the diabetic patient is required to be infused with artificial insulin using medicines and injections. Whereas, in type 2, the body gland (pancreas) produces insulin but is inadequate for the body. Majorly, type 2 diabetes is more prevalent among patients across the world. The population that is generally found to be affected by it comprise of adults, particularly people, in the middle-aged group, however, with the changing lifestyle and dynamics, its prevalence is found in children also.

The low prevalence of Type 1 diabetes is due to its interconnectivity with the external environment, which affects the body insulin-releasing cells. Though, the change in the lifestyle such as regular exercise and maintenance of adequate body weight can help prevent diabetes of type 2.

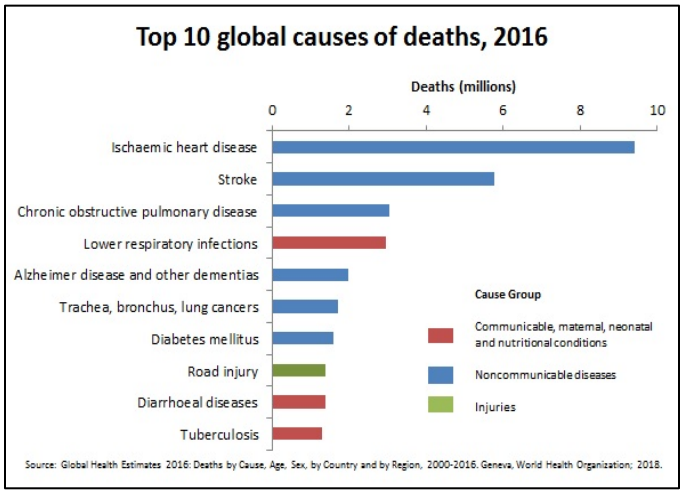


Figure 1: Primary Ten Causes of Death as Reported by WHO [1]

International Diabetes Federation has listed down following practices for preventing against diabetes:

1. It recommends that individuals must start a regime for losing some weight to overcome the effects of diabetes by enhancing insulin resistance and mitigating the prospects of hypertension. So, the people who are overweight are encouraged to sustain adequate body weight.
2. The consistency of physical exercise is integral for sustaining weight loss. It is because the indulgence in the physical activity overcomes arrhythmia, hypertension, and insulin sensitivity; improving the composition of the body; and developing psychological health.
3. Sustaining a healthy diet also overcomes the CVD related risk factors.
4. Smoking, depression, stress, and inadequate sleep can also be considered risky behaviors; therefore, it is preferable to avoid them.

B. Cardiovascular Diseases

Cardiovascular diseases comprise a number of disorders that are related to the heart and blood vessels for example;

Coronary heart disease which is caused by the clogging and narrowing or blockage of the blood vessels which supply the blood to heart muscles.

Cerebrovascular disease is caused by the issues in the arteries, which can affect the supply of blood to the brain. Example of the cerebrovascular disease is a stroke.

The peripheral arterial disease occurs due to plaque (calcium, fats, and other substances) that build up in the arteries supplying the blood to the head, limbs, and the other organs.

Rheumatic heart disease is caused because of the streptococcal bacteria, which attacks the tissues of the body, particularly of the brain and the heart.

Congenital heart disease is caused because of the anomalies of structure and the malformations at the time of birth.

Deep vein thrombosis and pulmonary embolism are caused due to the clotting of blood in the veins of the leg, which might be transferred to the organs like the lungs.

Myocardial infarction (MI) is the term which is used in the medicine for the NCDs in common, which also known as the heart attack. Cerebrovascular accident, which is also known as stroke, is also a kind of NCD and is a serious one. These illnesses or disease occur due to the defects and the faults of the heart and the arteries. The arteries are the vessels for the pumping of the pure blood from the heart into the body. Lack of exercise, bad eating habits, and fat accumulation in the body lead the fat cells to get deposited in the arteries' inner walls. Misuse of alcohol, the use of irregular tobacco, habits of eating, hypertension, and a host of circumstances and conditions lead to the CVA or MI.

Building up of fat inside the blood vessels and the arteries leading to the gradual vessel clogging. Without any treatment, medical care, or the changes in habits of eating or the lifestyles, it can lead to complete blocking of the blood flow. The unmanaged glucose levels in the blood, physical inactivity, and obesity are very common in the population nowadays. Formerly, people had a walking habit, they used to work in the farms, and the field were heavily involved in labor physically, which is rapidly reducing nowadays. Man, in search of comfort and the material gains, has

invented a lot of devices and machines, which has reduced the physical activity of humans. The emergence of the fast-food culture is also a danger to health and body, which most of us fail to understand.

If once diabetes, hypertension or the cholesterol imbalance is diagnosed, it is the high time to make sure that the person has the information regarding the upcoming dangers which might lead to the NCDs. These intermediate-risk factors defined by WHO must be highlighted by the primary health center and the clinics.

It is a general knowledge for the people who are having awareness that to stop or to reduce the dependency on the tobacco or the alcohol, to limit the oily and the greasy food that contains high fat and reducing sugar and the salt intake can help them in reducing the chances and the risks which are linked with the NCDs and the CVDs specifically. When a possible danger is detected, the patient should seek medical help immediately and make his lifestyles better. Further, people must use the proper medication for the control or restrict the damage that the factors can cause.

In most of the cases, there are no clear and visible indications for the MI and the stroke as well. It happens all of a sudden, and are not aware of it. Most of the MIs occur when people are sleeping. When the people feel the discomfort, and the pain in the chest or an extreme pain shoots up in the area of the right shoulder or jaws or the elbow, it can be considered as a warning that it can lead to developing of MI or CVA. The only option in this scenario is immediate medical attention.

The common symptoms which are linked with the CVAs are the insensitivity or the numbness in legs, arms, or face as well and it can be on one complete side of the body. There can be disorientation of speech, and even the difficulty in the vision as well. Severe dizziness or hallucinations or the headaches can happen, and in the intense cases, the patients got fainted too.

The third world is mostly affected by the NCDs; it is populated by middle and the low-income groups. The records of WHO speaks for themselves only. There is the accessibility to the medical care for the rich people while the people with the low or middle-income groups, if we look at the profiles of the country, have no or little access to the primary health care. In the countries across Southern America, Asia, and Africa, to approach the medical center is costly, which causes the late detection of NCD for the poor people.

According to the fact sheet of WHO [1], 26% of deaths in Bahrain are caused due to the CVDs. These statistics highlight the possibility of 13% deaths between the ages of 30 to 70 years and is caused by four NCDs. The physicians should record the factors of risks for the CVDs in order to reduce the strokes or the heart attacks by getting the right system for the storage of medical record and the analysis of the data.

C. Diabetes and cardiovascular disease: double jeopardy

Diabetes mainly contributes to the CVDs as proved by the clinical trials and the situation of the people who have suffered from CVDs. Due to the defects of crucial organs like kidneys and the liver, it becomes more difficult to pump the blood for the heart. The defects or the failure of kidney, liver, and the pancreas cause the threatening amount of toxins to sustain in the bloodstream.

The world of medicine has the challenge to handle this double-edged sword. This is the crucial time for the health centers and the governments to look at these two killers; CVDs and diabetes. To aid in reducing the deaths which are caused due to CVDs, data mining methods must be applied to predict the accurate occurrence or the reoccurrence of the CVDs and diabetes.

2.2. Evolution

Organizations nowadays produce a substantial amount of data specific to the institute. For improving the practicability concerning the use of data, researches have introduced algorithms which allow micro-focus on the data and position it as per the super-specific requirements. This advanced the efforts for the development and creation of the machine language algorithms which are useful in the analysis of the different analysis types and formation of decision without or little human supervision. Thus, the evolution of data mining is based on human needs, which assists in the identification of the relationship patterns and forecasting based on the presented layout of the program rules as well as stipulations.

The data mining is described by the researchers and practitioners using various terms. The concept of discovering knowledge from databases has been evolved from data mining. Earlier research of Fayyad et al.[4] has defined data mining as a procedure in which the data sets are implicitly and which reveal previously unidentified but significant information for effective decision-making. This whole procedure is termed as knowledge discovery in database (KDD).

This procedure can be applied in health care to predict the trends of many kinds of diseases and illnesses. Hence, instead of relying on the knowledge and experience, the data mining technique can be used by the doctors, more precisely for KDD to predict trends that would lead to better diagnoses.

KDD increases the efficiency and effectiveness of doctors by allowing them to treat a large number of patients at a given period. Moreover, such a system increases the opportunity for doctors of the same specialization across multiple firms, locations, as well as countries to e-share medical reports for devising best possible diagnoses in a time-effective manner.

The significance and usefulness of the data are evident from different perspectives. The logical alignment of the irrelevant data can be emphasized on the concealed or undiscovered correlations as well as patterns. This can provide valuable data which is critical for examining the individual as well as health being. The main notion is to briefly summarize the voluminous data and conclude its useful findings and information [5].

Data mining is regarded as a statistical interface, which is inclusive of other interference such as statistics, technology database, pattern recognition, data in machine-readable form as well as intelligent expert system [6].

Several definitions are set for data mining, which are raised according to the area of implication.

According to Krishnaiah et al. [7], the facility of data mining enables the use of data for identifying and using the data set trends. The primary findings of this database are to identify the

mechanical or automatic patterns, which require less input as well as efforts from the user.

Recently, the spectrum of data mining has enhanced, which is inclusive of artificial intelligence concepts which help in effective and fast-paced management and visualization of the data [8]. The other definition of data mining is provided by Han et al. [9], which states that it helps in the effective extraction of useful data.

Generally, the actual task of data mining is linked with the mechanical analysis of the voluminous amount of data, that is used to attain information which is not yet discovered. It assists in the identification of the data patterns, its categorization using the cluster analysis, odd records identification which require anomalies detection as well as its associated mining rule or dependencies.

Different statisticians' information system communities, as well as data analysts, use the data mining term. The procedure of KDD is recognized as complete, which requires the attainment of the data or discovery of new information. KDD core concept is of data mining; hence, it can be defined as an application which uses specific algorithms required for the analysis and extraction of the data patterns. It is also recognized as the KDD focal hub. It is inclusive of the intellectual approaches required for the data patterns extraction. From a healthcare perspective, the traditional methods are considered to integrate into statistical procedures for the process control comprising of numerous functions pertaining to the fundamental probability distribution which can be executed successfully for controlling the infection rate in hospitals [10].

Previously, in the 1960s, data mining was perceived to be an looked down by mathematicians as well as to statisticians alike. It was regarded as an unhealthy practice which improved its recognition using the term data fishing and data dredging, based on the hypothetical analysis of the data.

The terms database mining emerged in 1980 by HNC, a San Diego–based company, for describing the Database Mining Workstation [11]. With increasing years such as 1990, the term data mining was appeared to have been accepted as a legitimate phrase for describing the harvesting methods employed at the available data for forecasting the happenings in the future. Actually, different words were used for data mining, such as information harvesting, data archeology, knowledge extraction, information discovery, and so on [12].

The term "knowledge discovery in databases or KDD" was first initiated by Gregory Piatetsky-Shapiro and Parker [13]. Though, the relevance of data mining attained relevance as well as acceptance from the communities concerning artificial intelligence and machine learning. Furthermore, this term was also recognized within the business domain for the fourth estate communities. At present, both the terms, such as "data mining" and "knowledge discovery" are used interchangeably. Later in 2011, data mining was defined as data science.

2.3. Knowledge Discovery in Databases (KDD) Process

Knowledge discovery in databases (KDD) serves as a useful process that is used to extract important information from the expanded data. The information is gathered and filtered to extract only the required information. Information that has been collected is employed in the data set, which is further interpreted to attain a useful understanding regarding the given outcomes.

According to Maimon and Rokach [14], the target goals serve as important factors to successfully employ the KDD process. The process begins through fir, considering over the objectives, however, the final product is achieved in the form of newly developed information. This marks an end to the loop. Next involves data mining segments, where final output is achieved in the form of changes involved in the application domain. The results of the process are evaluated by employing fresh data sources while putting an end to the given loop. The finalizing of the process restarts the KDD process. Han et al. [9] elaborated the overall process through the given figure below:

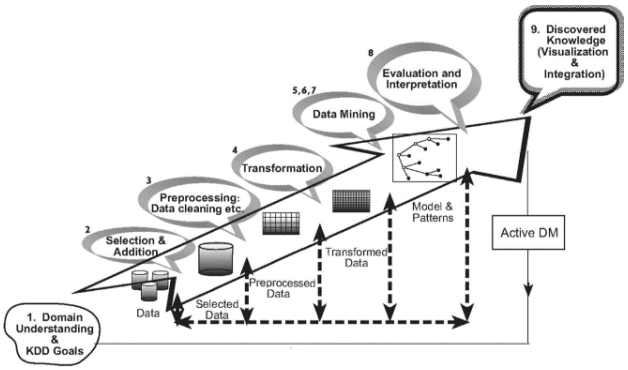


Figure 2. KDD process (Maimon & Rokach 2011)

Step 1. Develop a clear understanding of the application domains

The understanding related to the requirements of the healthcare professionals, and end-users, is of foremost value. This helps in providing a clear-cut vision to data miner in meeting the expectations while illustrating important knowledge regarding things that need explanation. It is further important to provide important knowledge regarding things that may help in achieving the desired objective. However, these requirements could be modified after the first round as the end-user might want to add additional functionalities. After identifying important knowledge, it is important for people contributing to the KDD process to function in accordance with the below-given steps. The steps are important for indicating preprocessing measures in the KDD process.

Step 2. Select data sets.

Selection of data sets is important to elaborate on the idea of the target goal. It further helps in providing important details regarding the type and the amount of data that is needed to achieve the target goal. The stage is of greater importance, since any wrong selection of data may lead to several complexities in preparing data. In such cases, the process may result in providing unimportant inferences that may weaken the overall effectiveness of the process.

Step 3. Perform preprocessing and cleansing.

Before processing the selected set of data, it must be filtered and cleaned. The idea is to improve and augment the reliability of the chosen set of data. The factor is important in providing maximum reliability to the selected data set. The cleansing

process includes removing noise or barriers that may impede processing; hence, eradicating the given barriers are crucial here. This might serve as a time-consuming process; still, it is effective in indicating maximum surety in terms of validity and reliability of data. In certain cases, insufficient data sets are achieved. However, situations, where ineffective data set, is intervened with accurate predictions, conscious efforts are required to attain expected outcomes. Consider an example; where irrelevant data of patient may act as a barrier in the overall process. Therefore, it is important to remove any such data at this stage

Step 4. Complete data transformation.

The next step is to make the data project-specific. This step ensures the provision of accurate data in a format that produces the required outcomes or results that will be utilized by the doctors to provide accurate predictions. The step is of significant value to provide valuable outcomes in any KDD project. However, important provisions in this regard may help in restoring the KDD process. Two significant methods, including; attribute transformation and dimension reduction are suggested by authors. KDD process serves as a useful tool in providing useful transformations. This helps in indicating maximum validity and efficacy to the formulated results.

Step 5. Choose the adequate data mining task.

Following the above procedure, the next task is to select the data mining type as per the nature of the project. This selection is based on the expected outcomes. Moreover, the selection between regression, and classifications or clustering is based on the desired outcome. Normally, the data mining outcome can be either forecasted, where data mining is done under controls of supervision, or description, where data mining occurs under visualization.

Step 6. Choosing the right algorithm for data mining.

This stage is reflected as the one which discusses the tactics to be used for obtaining the strategic objectives. Is it sufficient to use neural networks? Or is it better to use decisions? In fact, the selection is based on the searching pattern type which is found consistent to the given project and the desired end result. Does it relate to the results precision or decipherability or understandability that forms the project objective? The preference is more towards the neural networks when the precision of the results is required, while the appropriateness of the decision tree is found when gaining an understanding of the patterns and trends is required.

Step 7. Employ the algorithm of data mining.

In this step, the algorithm of data mining is implemented. It could be applied several times for obtaining the best result. For instance, the algorithm could be iterated using different controlling parameters, i.e., the low number of occurrences in a certain leaf or probably executed until the desired accuracy is obtained.

Step 8. Evaluate the mined data.

The mined data evaluation is the core of this KDD stage. This is primarily related to the reasonable interpretation and findings with respect to the discussed and defined goals of the project. This evaluation may promote the need to either add or remove a certain feature from the transformational stage of the data. This stage helps in achieving the usefulness of the data and its comprehension by the end user. Upon finalization, the found information is documented. When the found knowledge is being finalized by the KDD team, the whole data mining procedure is documented and assessed against the predetermined outcomes.

Step 9. Use of the discovered knowledge.

At this stage, the developed KDD process is evaluated for its effectiveness and efficiency by the end user, where refinement occurs in case the fine tuning of the data is required. Also, the created knowledge is being assessed for pilot-testing by practitioners and doctors for ascertaining the production of the desired outcomes. When a new product is being tested in a real environment, some conditions might variate from that at the laboratory. Therefore, these must comprise of the built-in ability to change, adapt, modify the derived knowledge in order to satisfy the end user.

2.4. Data Mining Techniques

Currently, an increased inclination of the researchers is concentrated on data mining. Generally, there are two types of data mining; namely, predictive model and the descriptive model. The two models are as follows;

A. Predictive Model

Prediction, as the name indicates constitutes of the correct envision of the future trend be logical computation of the data. The predictive model uses the previously available information for predicting future outcomes. This model is employed by various firms such as organizations who attempt to data mine the worth of an individual.

The predictive model data mining applies several techniques encompassing regression, classification, time series analysis, and prediction. This model is used for identifying the model that effectively matches the identified ideas or data sets. It is also helpful for class prediction of the objects when there is no availability of the class objects. The achieved model is primarily focused on the assessment of the identification classes set. To assess the numerical values and forecast, the statistical model of regression is used.

B. Descriptive model

This data mining model is employed to identify the data patterns to understand the relationship between the data attributes. The fundamental feature of the data is represented and summarized using the descriptive model. For instance, the customer database and identification of different sets can be useful to the marketers. The identified method can be employed for devising effective marketing programs for targeting audience.

The descriptive model applies several techniques such as clustering, summarization, association rules, and sequence discovery. In the clustering method, the analysis of the data object occurs without reference to the detected class label. Whereas, in the summarization, different properties and specialties of data values are to be considered as noise or outliers.

For the association rules, the same recreated rules are used for assessing the patterns and association among the different obtained characteristics. These are used for analyzing the data patterns for determining or forecasting their classification. Other systematic order identification is through sequence discovery which occurs for identifying the events in a timely manner rules and regulations. Figure 3 demonstrates the classifications of data mining.

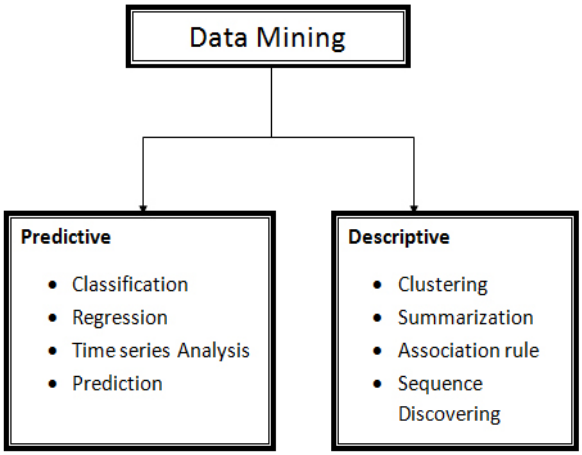


Figure 3. Tasks of Data mining

Among all the models presented in figure 3, the focus of the present study is on one named prediction-mining method.

2.5. Medical Applications Data Mining, Forecasting

Data mining has been widely used in medical applications to solve varies aspects. In this section we highlight some of them.

A. Health care informatics

Medical informatics is termed as dynamic because of the increasingly developing and shifting nature of both medical science and technology. According to Hersh and Hoyt [15], the discipline of health care informatics is related to the resources, materials, tools, and formal ways that are utilized for optimizing the storage, its management and retrieval of the needed biomedical information for effective decision-making, and problem-solving. Likewise, Athina and Konstantinos [16] highlighted that health care informatics employ modern communication, computer applications, as well as IT and computer system in medicine fields in the form of care provided, the delivered education, and research undertaken. Bronzino [17] stated that health care informatics uses different type of tool for using diverse which assists in utilizing and sharing information for improving the delivery of health care. Another term used is medical informatics whereas, it is also sometime referred by practitioners as clinical informatics or bioinformatics. Though, bioinformatics regard it as biological information analysis of the databases that are based on computer and statistical analysis of biological information in a speedy and improved manner. Clinical informatics concerns with maintaining data structure, data organization, and data search to make decisions using relevant data to conduct significant research in the medical discipline.

B. Data mining in medical discipline

Over the years, there is an increase in the computation analysis for deriving computer-based analysis and interference for improving the medical outcomes. The use of software and highly complex computation tools have increased. Previously, healthcare professionals majorly relied on their knowledge, skills, intuition and experience. However, at present, various channels have been developed across firms, regions, as well as states using both formal and informal sharing means for clinical experience which can be employed by the rest of the world. Data mining has provided great benefit to the medical field such as to forecast NCD and major other diseases.

This proliferation of the data mining has increased its efficiency in the medical discipline, where the use of data mining has become a norm where healthcare firms and facilities are being encouraged for meeting the need to predict the trends in disease and their occurrence. The significance of data mining, particularly for information management, is well established. Acharya and Yu [18] the developers of various computer-aided algorithms, have made the health facility more effective and have tremendously improved health care outcomes. Such as, it has aided in informatics, monitoring systems, as well as epidemiology. The difference between the pathological and normal data has facilitated improvement in the diagnosis and decision making.

Recent data mining techniques, as well as the essential statistical tools, can be used to assist doctors in diagnosing diabetes and CVDs. With the use of the statistical analysis, various CVDs encompassing stroke (cerebrovascular disease), cardiac arrests (coronary heart disease), congenital heart diseases, hypertension, peripheral artery disorder, arrhythmia, and heart failure with diabetes can be predicted. An overview of such systems is listed below.

Khaing [19] proposed an efficient approach for assessing and predicting the risk of heart disease. The primary stage is the execution of the k-means clustering algorithm with K=2 for extracting relevant data through clustering the heart disease database. By using k-parameter the number of fragments is mastered. Consequently, the extracted data of heart disease mine the frequent patterns with the use of maximal frequent itemset algorithm (MAFIA). Afterward, the execution of ID3 training algorithm occurs to produce the heart attack level following the decision tree. Although the accuracy of the findings obtained was 74% using k-mean–based MAFIA, there is a need to further increase this accuracy.

Many real-world medical data sets suffer from overlapping information. Overlapping k-means (OKM) is extended from the conventionally used k-means algorithm, and it is recognized as one of the most real-world medical data sets that inherent overlapping information. This clustering method enables one sample to be related to one or more than one cluster. Though, the issue of sensitivity in OKM is prevalent to the initial cluster centroids. Khanmohammadi et al. [20] proposed a hybrid method which assimilates the k-harmonic means and overlapping k-means algorithms (KHM-OKM). In this approach, the initialization of the cluster centers of OKM are obtained from the output of KHM. The performance of KHM-OKM is evaluated in

terms of completeness, homogeneity, and cluster size-quantity tradeoff, and it is found that it outperforms the OKM algorithm.

Feature selection is one of the integral steps in data mining. When selected properly, it improves prediction accuracy. Orthogonal Feature Extraction (OFE) is proposed by Jiang et al. [21], which uses the feature ranking techniques. The technique proposed is applied to improve cancer prediction accuracy. To make sure that the attributes of the selected features are accurate, they developed an algorithm that selects the linearly independent vectors iteratively that belong to top-ranked attributes. The results are compared with the analysis of principal component, neighborhood component, and linear discriminant. These three methods are outperformed by OFE in terms of computational complexity and performance.

Another heart disease prediction system is developed by Suvarna et al. [22]. This system combines optimization techniques and data mining. The optimization technique used is particle swarm optimization, which is modified by applying a constriction factor. In this inherently distributed algorithm, the solution is obtained through the interaction among several simple individuals called particles. The authors used *de facto* standard data sets for the reliability ranking of heart disease prediction. The three techniques used for comparing the result are: ID3, multilayer perceptron with backpropagation training and classification, and regression trees. The proposed approach outperforms these three techniques, producing accuracy of classification equal to 53.1%. But the accuracy is much lower than what is expected from actual practitioners.

Rairikar et al. [23] applied the backpropagation technique in genetic algorithm to develop a system that predicts. This prediction system uses 13 attributes obtained from cardiovascular disease information. The results are compared in terms of time complexity with two well-known classification of data mining techniques, which are KNN and Naive Bayes. The results showed that KNN is much better in its performance. However, this method is lagging measuring the accuracy of heart disease prediction.

Tomczak and Zieba [24] applied classification data mining in machine learning to solve two issues that appear when applying machine learning to medical diagnosis. One issue deal with the implementation of interpretable models such as decision tree and classification rules. The other issue deals with the imbalance between classes with high number of examples such as healthy patients and a class with low number of examples such as ill patients. The proposed model is a probabilistic combination of *soft rules*. These rules are constructed by introducing a new random variable called *conjunctive feature*. In addition, a new estimator is introduced to incorporate the knowledge about imbalanced data. This approach is tested using oncology data set. The results show that soft rules can perform well on data sets with small number of examples. In addition, the outcome is comparable with expert oncologists. Meanwhile, this work needs to be examined against large data sets.

Zhu and Fang [25] noticed that the current existing classification trees might not be able to properly provide classification for the provided predictor variables set, which may be categorized using a high error rate. To solve this, they proposed an algorithm that combines the logistic regression model and the trichotomous classification tree. This tree is used to split and establish the tree recursively until it meets the stopping rule of the tree splitting. The algorithm is tested on two real data sets: Pima Indian data set and Wisconsin Breast Cancer data set. The findings showed that the proposed algorithm is more accurate in predicting some diseases than the classification and regression tree. However, this technique suffers several limitations, including using binary response variable, the cost of variables used in classifications ignores the actual cost and thirdly, the variance-covariance between 2 categories use limited simulated normal distributions.

Several algorithms and methods such as regression, clustering, classification, neural networks, artificial intelligence, genetic algorithm, association rules, nearest neighbor, and decision tree have been executed by researchers to assist healthcare practitioners in the effective diagnosis of heart diseases or diabetes to prevent further diseases.

The above-mentioned research studies provide evidence for the effectiveness of the data mining concerning the field of NCDs prediction and correction. The objective of this research is to devise a practical model which can be executed by the healthcare practitioners in the main hospital of Bahrain, or where individuals belonging to particular demographics which adds to its generalizability across the worldwide health care centers.

What features the developed PNCDA is that it allows realistic and timely mean to forecast the general NCD forms, particularly CVDs and diabetes. Rather than using pre-defined datasets, this approach uses real data obtained form BDFH. Further, the developed PNCDA has been tested by actual practitioners from the mentioned hospital and the results met their expectation obtained from the conducted questioner as will be explained later.

3. Research Methodology

As the major outcome of this research is the prediction software system that aids the doctors and practitioners of the BDFH to effectively predict the NCDs of their patients, the first stage of this project is to identify the most effective factors and classify them. These factors must provide good input data (parameter) using an adequate model for data mining. These factors can be identified by the users of the proposed system. Three main factors are included in this research, such as, prediction methods for data mining, patient database, and software application implementation, which are used to characterize and devise a model to predict the cardiac arrest diagnosis as well as diabetes risk among the patients that are listed in the obtained BDFH data set. This research will use both forms of data (primary and secondary). It will integrate in data mining, its significance in healthcare, and medical field application for developing a software system that will be assessed in real situations for evaluating its performance.

3.1 Data Types

In this research, two types of data are used; secondary data and primary data.

A. Secondary data

Secondary data refers to the data that has already been assessed, used and collected by certain firms. In this research, the record of 23000 patients is obtained from BDFH, which are used for designing the required NCD prediction application.

B. Primary data

The data that is collected for the first time as per the defined objectives of the research are termed as primary data. The primary data for the study is collected using a questionnaire which is distributed across 30 specialized healthcare professionals in the hospital. Since all the questionnaire were completed and received; therefore, the response rate is termed as 100 percent. The objective of the questionnaire is to gather the necessary requirement, NCD diagnosis for medical characteristics, which include CVD and diabetes, which are the commonly found forms of NCD.

3.2 Data Collection Procedure

A survey was conducted using the questionnaire as a primary form of data collection. While, for the secondary form of data collection, the health center record was used following the development of the software application.

A. Primary data collection

The questionnaire-based survey was conducted among the doctors to clearly understand the expectations they hold as well as ideas about the current situation concerning NCDs from one side and to examine the awareness and the level of utilization of having a new software application that will help them in predicting NCDs from the other side. To evaluate the current situation concerning NCDs prediction, the BDFH doctors who are either endocrinologists or cardiologists were requested to fill in a questionnaire survey. It was done to gain an understanding pertaining to their perceptions and opinions about the expected software application.

B. Study Population and Sampling

In BDFH, 13 cardiologists and 17 endocrinologists are present. The sample of the study is the entire population given the relative ease to work around. Therefore, the sample would constitute of all the related doctors. The structured questionnaires were administered by the researcher to assess their perception and opinion.

4. Survey Data Analysis and Findings

This section provides information concerning the questionnaire model, which was used in the survey. The survey aimed to establish the importance of predicting model concerning NCD trends among the end-user, and doctors. The direct interviews were held with the doctors to explain to them the significance of developing the software.

The conducted questionnaire was aimed to examine the present scenario concerning the use of NCDs forecasting tools in the BDFH. The sample is formed by the total endocrinologists and cardiologists' populations who are presently on the rolls of the BDFH. As the interviews are conducted on a one-to-one basis; therefore, no interview was missed. The following results were achieved by the questionnaire;

The answers to the first question show that 70% of patients treated by these 30 doctors suffer CVDs and diabetes. This provides evidence that NCDs, more particularly, diabetes and cardiovascular disease are majorly found. The generic and common disease such as influenza cold, allergies, and sprain make a total of about 30 percent. The BDFH does not provide treatment for critical diseases such as cancer.

The second question examines the primary NCDs causes, where the doctors mentioned unhealthy eating habits, smoking, no exercise, and consumption of high-fat food as the major NCDs reasons.

Following third and fourth questions in the questionnaire are related to each other and are also analyzed together. Though 80 percent of the doctors are aware of the prevalence of some form of NCD prediction mechanism across the globe, they firmly say (i.e., 100%) that BDFH lacks any such system, mechanism or facility.

In the fifth and sixth questions, 24 among the 30 doctors (i.e., 80%) showed agreement that presence of computer-based application would assist in predicting such NCDs, which aids the BDFH in notifying patients on a continuous basis. In addition, 90 percent of them welcomed the idea of offering an application that can "correctly predict" the NCDs occurrence, which is essential.

5. Proposed Application System

The software system developed comprises of two main components; the prediction engine and the software application. The system functions as an intermediator (practitioner) between the prediction system and the user. The explanation of each component is presented in which first describes the data that is applied in the proposed prediction system.

5.1 Data Mining Engine

The application purpose is to assess the data of the patients to predict if the patient is at an NCD risk or not. The following tools are used for application programming;

1. VS.Net: Tool for writing the application.
2. VB.Net: Tool for coding the application.
3. Oracle DB: The database which is used for data storage and manipulation.
4. Toad Oracle: Tool for gathering the code and retrieving the data from the database.

Among various techniques for data mining, the research adopted the mining class comparison. This is used to mine a description which compares or distinguishes one class from another related class. It is used for meeting the project purpose based on stages which are followed for the process of mining comparison [9]. The stages are as follows;

1. *Data collection*. It comprises a set of relevant and useful data that is collected by the use of query processing. The segregation of the data occurs in the target class along with contracted classes set.

2. *Dimension relevance analysis*. This is executed only where there are several dimensions and only a few relevant dimensions are to be selected for analysis.
3. *Synchronous generalization*. This is applicable on the target class concerning the degree of control exercised by the user. The contrasting classes concept is to generalize the same level as those in the main target relation, to form the prime contrasting classes' relation.
4. *Presentation of the derived comparison*. The achieved class comparison can be represented in the form of tables, graph, and rules.

5.2 Patient Data Set

The researches, which tackle data mining for medical application use the available data sets such as heart-Statlog Medical data set [26], liver disorder, Indian liver patients, heart disease (Statlog), heart disease (original), hepatitis Parkinson's, Parkinson's, breast cancer Wisconsin (diagnostic), dermatology, and lung cancer [20].

This work is featured using the actual data set obtained from the BDFH, which contains the details and vitals of 23,000 patients.

Using the developed data set, that was extracted with the use of the discussed algorithm of data mining, the recognized pattern assists to predict the risk of getting heart disease and diabetes.

The data of all the patients were processed through the elimination of the repetitive records and completion of the missing details. The dummy values were used for filling the personal details (name and ID number, etc.). Moreover, the recurring patterns are mined with the use of proposed process of data mining. The attributes of the patients gathered from the database are; sex, age, chest pain (CP), FBS (fasting blood sugar), test BPS, rest ECG (electronic cardiographic), smoking, diet, as well as alcohol. Following it, these are used for classifying the patients, which are modified from the study conducted by Khaing [19] as some of them are unavailable in the data set, we have. Afterward, they are adjusted as per the survey conducted with the actual practitioners in the BDFH.

5.3 Application Process Description

The overall exercise was conducted with doctors who will be the main user of the developed software. This is to ascertain that all the unnecessary data is removed so that the prediction of NCDs is barrier-free. The data of the patients is derived from the three tables by executing the second step of KDD. Figure 4 presents the diagram for the entity relationship (ER). The information record of the patients is presented in the first table, second patients' vitals, such as, diet, blood pressure, chest pain type, smoking, and heart rate. The blood test collected from the lab is collected in table 3, where cholesterol and fasting blood sugar are its attributes.

Reflecting that the target class are the lab results and vitals, the "if condition" is used for the constructing class, along with attributes such as sex, age, chest pain (CP), FBS (fasting blood sugar), test BPS, rest ECG (electronic cardio graphic), smoking, diet, as well as alcohol (Table 1). The data mining query language (DMQL) can be used for the data mining task, as follows;

1. Used Patient_DB
2. Mine comparison as "Patient_NCD_Risk_Level."
3. In relevant to sex, age, CP, chol, FBS, test BPS, rest ECG, smoking, diet, alcohol
4. For "Vitals_View" and "Lab_Results"
5. Versus "if_condition_rule"
6. Evaluate sum
7. Show as "risk level."

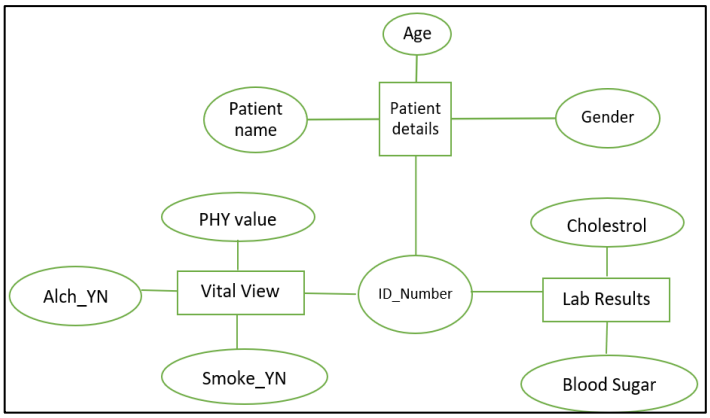


Figure 4. ER diagram of the proposed application

Initially, the change in the query occurs in two forms such as in the two interrelated queries where the two sets of relevant data tasks are accumulating; such as, one is obtained from Vital View table target class and Lab Results table, whereas the other is derived from the contrasting class, i.e., "if condition rule" which is written in the code classes.

Secondly, the two tables, i.e., Vitals View and Lab Results, are analyzed for their dimension relevancy. Furthermore, the weakly relevant dimension are removed, i.e., the removal of the patient information from the found test classes.

Thirdly, the synchronous generalization is applied from the target class to the controlling class level in order to produce the prime relation for the target class. With the use of "if condition" rule, the categorization of the attributes in the form of group is held. Initially, the ages are evaluated. In case, the age is less than 40, it is grouped as young age, whereas, if it lied between 40 to 60, it is grouped as medium age while for 60 or greater than 60, it is regarded as older age. Following it, smokers or alcohol drinkers are categorized. In case the patient is a smoker, he is characterized as group 1, and in group 2 if the situation is vice versa. Similarly, if the patient is an alcohol drinker then he is characterized in group 1, and in group 2 if he does not.

If the blood pressure (BP) of the patient is less than 80, then he is grouped into group 0, and group 1 if it is equal to or more than 80 and group 2 when the blood pressure is higher than 90. Likewise, for the classification of heart rate, if the heart rate is below or equal to 100, it is categorized in group 0 while if it is more than 100 and less then 150, it is categorized as 1. Similarly, in case the heart rate is greater or equal to 150, the patients are categorized in group 2.

Table 1. Attributes and values used in the proposed application

No.	Attributes	Description	Cut-off Value	Type
1	Age	Age of the	Age <=40 Age <=60 and >40	Young age Middle age Old age

		patients in years	Age >60	
2	Gender	Gender	1 0	Male Female
3	CP	Type of Chest Pain	Value 1 Value 2 Value 3 Value 4	Stable Angina Unstable Angina Nonangina Pain Asymptomatic
4	Test BPS	Resting blood pressure (in mmHg)	BP <80 BP <90 BP >90	Normal Normal to high High
5	Chol	Serum Cholesterol (mmg/dl)	Chol <5.2 Chol <5.2 and >6.2 Chol >=6.2	Normal High Severe
6	FBS	Fasting blood sugar	1 0	True False
7	Rest ECG	Resting electro cardiographic results	Val=0 Val=1 Val-2	Normal Abnormal Probable
8	Diet	On a healthy diet	0 1	True False
9	Smoking	Smoker patient	0 1	True False
10	Alcohol	Alcohol drinker	0 1	True False

```
If txtAge.Text.Trim() <= "40" Then
    strAge = "0"
    'MessageBox.Show("Young Age")
ElseIf txtAge.Text.Trim()> "40" And txtAge.Text.Trim() <= "60" Then
    strAge = "1"
    'MessageBox.Show("Middle Age")
ElseIf txtAge.Text.Trim() > "60" Then
    'MessageBox.Show("Old Age")
     strAge = "2"
End If
```

Figure 5. Comparison of condition for the age

Lastly, Table 1 presents a comparison of the resulting classes in the rules form. Such as, in case the classes are equal to 0 when there is no risk for patients. While patients are at high risk when the classes are equal to 2. Likewise, the patient is at lower risk, when the classes are equal to one, where the other classes combination occurs based on the recommendation of the actual practitioners.

Following the preprocessing, the previously mentioned attributes in Table 1 are utilized for the code application for using the IF condition. The classification of these values done in the form of cut-off value, following its comparison.

```
If Trim("" & sqlds.Tables(0).Rows(0).Item("SMOKE_YN")) = "Y" Then
       cmbSmoking.SelectedIndex = 1
ElseIf Trim("" & sqlds.Tables(0).Rows(0).Item("SMOKE_YN")) = "N" Then
       cmbSmoking.SelectedIndex = 2
End If

If Trim("" & sqlds.Tables(0).Rows(0).Item("ALCO_YN")) = "Y" Then
       cmbAlco.SelectedIndex = 1
ElseIf Trim("" & sqlds.Tables(0).Rows(0).Item("ALCO_YN")) = "N" Then
       cmbAlco.SelectedIndex = 2
End If
```

Figure 6. Grouping smoke and alcohol

The process of grouping the attributes are described by explaining the Figures 6–10 as follows:

Figure 6 shows the comparison if patients smoke or drink alcohol. In case the patient is a smoker, he is characterized as group 1, and in group 2 if the situation is vice versa. The same situation is followed if the patients is an alcohol drinker than he is characterized in group 1, and in group 2 if he does not.

Figure 7 presents the BP, for which if the patient BP is less than 80, then he is grouped into group 0, and group 1 if it is equal to or more than 80 and group 2 when the blood pressure is higher than 90.

```
If txtBP.Text.Trim().Substring(4, 2) < "80" Then
       strBp = "0"

ElseIf txtBP.Text.Trim().Substring(4, 2)>= "80"
      And txtBP.Text.Trim().Substring(4, 2) <= "90" Then
       strBp = "1"

ElseIf txtBP.Text.Trim().Substring(4, 2) > "90" Then
       strBp = "2"
End If
```

Figure 7. Grouping blood pressure (BP)

Figure 8 demonstrates the grouping of heart rate. If the heart rate is below or equal to 100, it is categorized in group 0 while if it is more than 100 or less then 150, it is categorized as 1. Similarly, in case the heart rate is greater or equal to 150, the patients are categorized in group 2.

Figure 9 displays the patients grouping as per his cholesterol level. The patient is grouped in 0, when the level of cholesterol is below 5.2, and in group 1 if the value of cholesterol is more or equal to 5.2. Similarly, these are grouped 2, when it is more than or equal to 6.2.

Figure 10 provides a list of conditions for chest pain. As per the program, if the grouping of the chest pain is done as 1, the patient experience chest pain, and 2 in case a risk of cardiac arrest prevails. It is grouped as 3 if a patient is experiencing coronary artery disease. Whereas for being grouped as 4, the patient is at high risk of chest pain.

Lastly, the class comparison results are presented in rules forms (Table 2) and process flow of the system is depicted in Figure 11.

5.4 Testing Results of the Application and its Predictive Performance

As explained above, the resulting conditions provide a comparison of the prime target classes and the executed "if condition rule" (Table 2), where risk results are the outcome of applying the aforementioned rules.

```
If txtHeartRate.Text <= "100" Then
       strHR = "0"

ElseIf txtHeartRate.Text > "100" And txtHeartRate.Text < "150" Then
       strHR = "1"

ElseIf txtHeartRate.Text >= "150" Then
       strHR = "2"

End If
```

Figure 8. Grouping heart rate

```
If txtCholesterol.Text < "5.2" Then
       strTotCh = "0"

ElseIf txtCholesterol.Text >= "5.2" And txtCholesterol.Text < "6.2" Then
       strTotCh = "1"

ElseIf txtCholesterol.Text >= "6.2" Then
       strTotCh = "2"

End If
```

Figure 9. Grouping cholesterol

```
If cmbChestPain.SelectedIndex > 0 Then

   If cmbChestPain.SelectedIndex = 1 Then
       strChPain = "1" 'Predictable Chest Pain

   ElseIf cmbChestPain.SelectedIndex = 2 Then
       strChPain = "2" 'Chest Pain Signal to Heart Attack

   ElseIf cmbChestPain.SelectedIndex = 3 Then
       strChPain = "3" 'Have Coronary Artery Disease

   ElseIf cmbChestPain.SelectedIndex = 4 Then
       strChPain = "4" 'Predictable Chest Pain

   End If

End If
```

Figure 10. Grouping chest pain

The software application is implemented using VB, and the samples of the screenshots are depicted in Figures 12 to 15. These screenshots serve as evidence for the program utility concerning the NCDs prediction for the tested patients. These samples of NCD predictions are shown for different risk levels: no risk, low risk, medium risk, and high risk.

5.5 Findings

The results are examined on the basis of the response gathered through a questionnaire survey, which was held among the doctors and the results achieved from the use of the developed application.

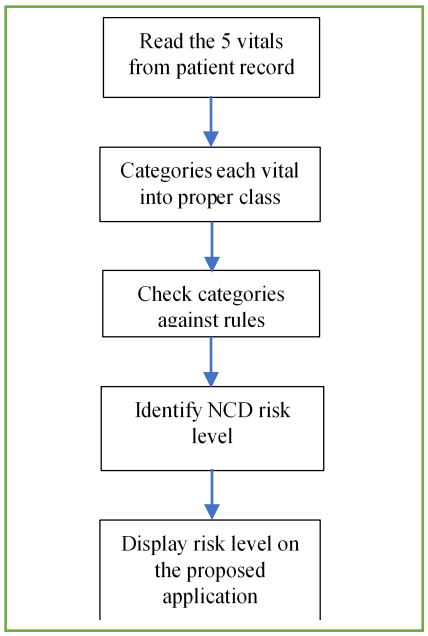


Figure 11. Flow diagram of the proposed application

Table 2. Classes and Risk Results Comparison

BP	Heart Rate	Cholesterol	Chest Pain	Age	Risk Result
0	0	0	0	0	No risk
0	1	0	1	0	No risk
0	2	0	1	0	No risk
1	1	1	1	1	Low risk
1	2	0	1	0	No risk
1	2	0	1	1	No risk
1	2	0	1	2	Low risk
1	2	0	2	2	Medium risk
1	2	1	1	1	Low risk
1	2	1	2	1	Medium risk
2	2	1	1	1	Low risk
2	2	1	1	2	Medium risk
2	2	1	2 or 3	2	High risk
2	2	2	1	1	Medium risk
2	2	2	2 or 3	2	High risk

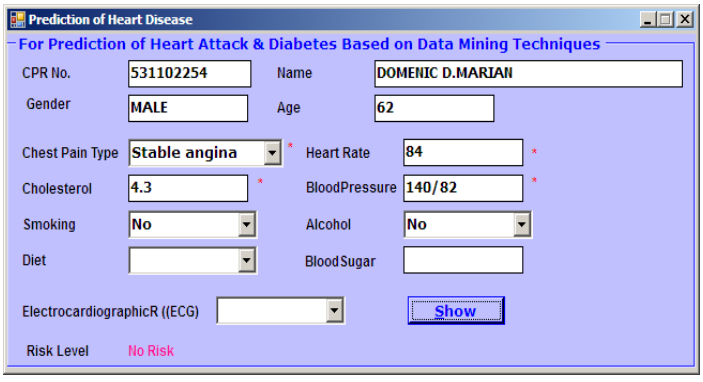


Figure 12. Sample of NCD prediction level: no risk

A. Survey Based Questionnaire Inferences

The questionnaire survey provides evidence that at present, there are no tools or the mechanism available in the BDFH, which is having almost 23,000 registered patients, for NCD forecasting

and prediction. The patient's record provides a massive volume data, yet there is no examination or the analysis of such essential data for the purpose of forecasting. The survey highlighted that most of the doctors are aware regarding the tools which are being used somewhere else and they will like to have applications that can help them better in the prediction of a patient's possibility for the occurrence or the recurrence of NCDs.

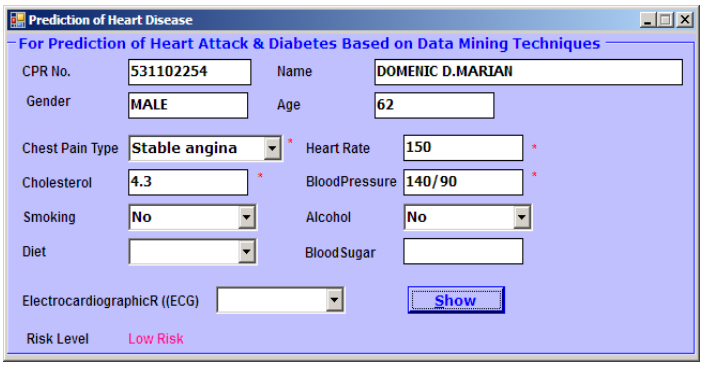


Figure 13. NCD prediction level (sample): low risk

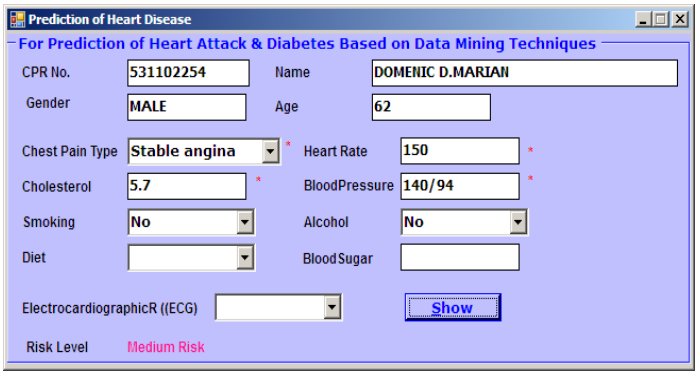


Figure 14. NCD prediction level (sample): medium risk

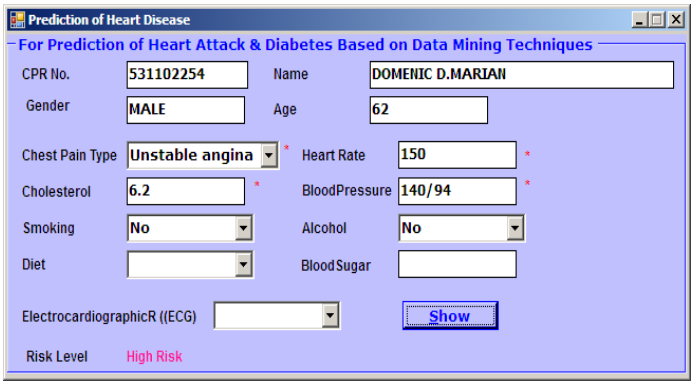


Figure 15. NCD prediction level (sample): high risk

B. Analysis of the developed software Inferences

The results obtained were discussed with the doctors and are considered as the levels of risks. The doctors were highly contented with the results as they accurately matched their expectations regarding the levels of risk. All kinds of risks are categorized appropriately according to the formerly mentioned categories. Additionally, they indicated high interests in the implementation of such applications live in their clinics. The application of the software has a proof for the accuracy and the efficacy of the suggested technique of the data mining for the forecasting of the NCDs for the patients of BDFH. However, the NCDs can lead to the possible mortality among the patients, the doctors who had a trial of the software feel excitement for its application as it can help them in the detection. As a matter of fact, they want to use it as soon as possible because they highly noticed that it supports quickly and gives better forecasting along with the briefing of the patient's record. On the other hand, the statistical analysis is not applicable to this study as the results are evaluated manually by the consulted practitioners.

Using the application, more savings in medical expenditure can occur avoiding loss of the duty timings and maximizing the utilization of the crucial medical facilities. This application will give the one-stop-shop to all the patients to get the accurate and the calculated information regarding their health. This application is considered to be an asset for the doctors so that they make sure that their detection or inferences are professional as well as correct. Further, such information can be utilized globally, for the patients having the health problems and have to travel abroad, the medical history and the forecasting for the NCD can be made online available or can be shared via emails.

6. Conclusions and Future Work

An application which uses the data mining algorithm for the class comparison has been invented to forecast the level of risk of occurrence or recurrence of NCDs such as diabetes and the heart diseases. Additionally, the outcomes of the application highlighted that the forecasting system could forecast NCDs efficiently, instantly, and effectively. This application helps the doctors to make the decisions regarding the health risks of a patient. It also creates the results, which make it closer to the situations of real life. Data mining, therefore, is more supportive for the health sector and is essential for exploring the knowledge to be used in the health care sector.

This research confirms that the health centers should have a software application or the system which can forecast the NCDs. To have raw data in a large volume is a must for the proper categorization rules which lead to the proper forecasting of NCDs. The forecasting is helpful for not only the physicians but for the patients too for their constant health check-ups. The application developed must be trialed on a constant basis, and the coding must be accurate in order to adopt the application across the various health centers and the clinics. Further, in the future, it can also be considered for the development of the mobile application for the patients where they can input their vitals and get the results instantly.

The developed application and the software can be extended for assisting in forecasting the level of risks for the other NCDs, for example, cancer, Alzheimer, chronic lung disease, stroke, and the osteoporosis. This can be achieved more by highlighting the associated vitals and their analysis to set the appropriate rules by the consultation linked with the expert practitioners. Finally, the obtained results can be further analyzed once the approach is applied on predefined data set with previously known level of risk of each NCD disease.

References

[1] Anon., 2016. World Health Day 2016. [Online] Available at: http://www.who.int/diabetes/global-report/WHD16-press-release-EN_3.pdf?ua=1.

[2] W. H. Organization, "The top 10 causes of death," 24 May 2018. [Online]. Available: http://fmrglobalhealth.com/frame/top10.html. [Accessed 28 June 2019].
[3] Anon., 2018. Diabetes country profiles. [Online] Available at: https://www.who.int/diabetes/country-profiles/bhr_en.pdf?ua=1 [Accessed 28 June 2019].
[4] Fayyad, U. M., Piatetsky-Shapiro, G. & Smyth, P., 1996. From data mining to knowledge discovery. In Advances in Knowledge Discovery and Data Mining, AAAI Press/MIT Press, pp. 495–515.
[5] Clifton, C., 2010. Encyclopaedia Britannica: definition of data mining.
[6] Obenshain, M. K., 2004. Application of data mining techniques to healthcare data. Infection Control and Hospital Epidemiology, 25(8), pp. 690–695. DOI: https://doi.org/10.1086/502460
[7] Krishnaiah, V., Narsimha, G. & Chandra, S., 2016. Heart disease prediction system using data mining techniques and intelligent fuzzy approach: a review. International Journal of Computer Applications, 136(2), pp. 43–51. DOI: 10.5120/ijca2016908409
[8] Yoo, I. et al., 2012. Data mining in healthcare and biomedicine: a survey of the Literature. Journal of Medical Systems, 36, pp. 2431–2448. DOI: https://doi.org/10.1007/s10916-011-9710-5
[9] Han, J., Pei, J. & Kamber, M., 2011. Data mining: concepts and techniques. Walthman, MA: Morgan Kaufman Publisher. DOI: 10.1007/978-1-4419-1482-6_3752
[10] Benneyan, J. C., Lloyd, R. & Plsek, P., 2003. Statistical process control as a tool for research and healthcare improvement. Quality & Safety in Health Care, 12, pp. 458–464. DOI: 10.1136/qhc.12.6.458
[11] Mena, J., 2011. Machine learning forensics for law enforcement, security and intelligence. In CRC Press.
[12] Campos, J. et al., 2017. A big data analytical architecture for the asset management, Elsevier, pp. 369–374. https://doi.org/10.1016/j.procir.2017.03.019
[13] Piatetsky-Shapiro, G. & Parker, G., 2011. Lesson: data mining, knowledge discovery: an introduction. In Introduction to Data Mining.
[14] Maimon, O. & Rokach, L., 2011. Data mining and knowledge discovery handbook, Springer.
[15] Hersh, W. R. & Hoyt, R. E., 2018. Health informatics: practical guide seventh edition. lulu.com.
[16] Athina, L. A. & Konstantinos, S. M., 2008. Handbook of research on distributed medical informatics and e-health. Medical Information Science Reference.
[17] Bronzino, D., 2006. Medical devices and systems, CRC Press.
[18] Acharya, R. U. & Yu, W., 2010. Data mining techniques in medical informatics. The Open Medical Informatics Journal, 4, pp. 20–21. DOI: 10.2174/1874431101004020021
[19] Khaing, H. W., 2011. Data mining based fragmentation and prediction of medical data. Shanghai, pp. 480–485. DOI**:** 10.1109/ICCRD.2011.5764179
[20] Khanmohammadi, S., Adibeig, N. & Shanehbandy, S., 2017. An improved overlapping k-means clustering method for medical applications. Expert Systems With Applications, 67, pp. 12–18. DOI: 10.1016/j.eswa.2016.09.025
[21] Jiang, H., Ching, W. K. & Hou, W., 2016. On orthogonal feature extraction model with applications in medical prognosis, Applied Mathematical Modelling. 40, pp. 8766–8776. DOI: 10.1016/j.apm.2016.05.011
[22] Suvarna, C., Sali, A. & Salmani, S., 2017. Efficient heart disease prediction system using optimization technique, Erode, India, pp. 374–279. DOI: 10.1109/ICCMC.2017.8282712
[23] Rairikar, A. et al., 2017. Heart disease prediction using data mining techniques, Coimbatore, India, pp. 1–8. DOI: 10.1109/I2C2.2017.8321771
[24] Tomczak, J. & Zieba, M., 2015. Probabilistic combination of classification rules and its application to medical diagnosis. Machine Learning, 101, p. 105–135. doi: 10.1007/s10994-015-5508-x
[25] Zhu, Y. & Fang, J., 2016. Logistic regression-based trichotomous classification tree and its application in medical diagnosis. Medical Decision Making, 36(8), pp. 973–989. DOI: 10.1177/0272989X15618658
[26] Jaganathan, P. & Kuppuchamy, R., 2013. Threshold fuzzy entropy based feature selection for medical database classification. Computers in Biology and Medicine, 43(12), pp. 2222–2229. DOI: 10.1016/j.compbiomed.2013.10.016

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 207-212 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Integrating Diacritics Restoration and Question Classification into Vietnamese Question Answering System

Bui Thanh Hung*

Data Analytics & Artificial Intelligence Laboratory, Engineering - Technology Faculty, Thu Dau Mot University, 590000, Vietnam

ARTICLE INFO

Article history:
Received: 30 May, 2019
Accepted: 13 September, 2019
Online: 08 October, 2019

Keywords:
Diacritics Restoration
Question Classification
Vietnamese Question Answering System
Deep Learning
Encoder-Decoder LSTM
Bi-LSTM

ABSTRACT

This paper presents a solution for question answering system for Vietnamese language by integrating diacritics restoration and question classification via deep learning approach. It could be said that this will be the first research integrating two phases into Vietnamese question answering system. Question classification has a critical role in the question answering system. However if the question has too many missing diacritics, this will make the classification extremely more difficult. In this paper, both automatic insertion of diacritics and question classification tasks are built to rely on deep learning approach. For diacritics restoration task, we apply the Encoder-Decoder LSTM model. The result of the first step will be the input of question classification. We use pre-train word embeddings in the Bidirectional LSTM model for Vietnamese question classification. The deep learning approach for both tasks is powerful and highly accurate model. By integrating diacritics restoration and question classification into Vietnamese question answering system – ICTbot of Binh Duong Department of Information and Communications Support System – it has produced remarkably positive results; thus proves the practicability of this proposed system.

1. Introduction

Internet allows the researchers to save data and make them available to the public. However, it also makes the search of information in such big data environment more complicated and expensive. Question Answering systems have been developed as a new advanced research tools for this issue.

An automated question answering system has several key steps to follow such as:

- The user inputs the question through query interface. The question is analyzed, pre-processed and phrased into words.
- Question classification identifies the type of the question. This step is rather critical to achieve the most appropriate answering type.
- Finally, based on the result of Question classification, the system will choose the appropriate answer from the answer sets.

In the first step, if there are too many missing diacritics in questions, it will make it challenging for the second step - Question classification. A diacritic is a mark written either above or below a letter to alter its original phonetic or orthographic value. Diacritics are used in several languages' orthography. Nevertheless, in daily conversation, as a matter of convenience, diacritics are usually ignored in many languages. The absence of diacritics in question text makes it extremely challenging for both Question Classification and Question Answering system.

The Vietnamese alphabet uses Latin script with diacritical signs. However, Vietnamese has seven modified letters including ă, â, đ, ê, ô, ơ, and ư and also six tone marks including unmarked tone (ngang), acute (sắc), grave (huyền), hook above (hỏi), tilde (ngã), and underdot (nặng). Due to its diacritical complexity, when typing on phones or computers, most of Vietnamese people prefer to type non-diacritic characters as a matter of convenience. As a consequence, it sometimes causes the confusion or mis-interpretation for the recipients or readers. With that being said, integrating diacritics restoration into question classification is a critical task for Vietnamese question answering system.

We propose the Vietnamese Question Answering model with two phases in this paper. The first phase is automatically recovering the missing diacritics. And the second phase is integrating diacritics restoration into Vietnamese question

*Bui Thanh Hung, Engineering - Technology Faculty, Thu Dau Mot University, hungbt.cntt@tdmu.edu.vn

answering system. We use the Encoder-Decoder LSTM model for automatically recovering the missing diacritics and the Bidirectional LSTM model for Vietnamese question classification. We integrate diacritics restoration into Question classification of Vietnamese question answering system.

This paper consists of five sections. The related works are reviewed in Section 2. Our method is presented in Section 3. Section 4 mentions the procedure of constructing dataset, the experiment settings and discusses the outcomes of the experiments. Our work and possible future research are concluded in Section 5.

2. Related Words

Firstly, diacritics restoration – as the first step – could be run following two typical approaches: character-based [1-2] and word-based [3-5]. For Vietnamese language, many researches propose their methods to restore accents such as [6-9]. In this research we choose the Encoder-Decoder LSTM model that proposed recently by Bui [6] because of the advanced approach, and high accuracy for Vietnamese diacritics restoration.

For question classification, there are three main approaches: Rule-based, Machine Learning and Hybrid ones [10]. Using a few manually handcrafted rules to match the question is the rule-based approach. One of key challenges here is to define too many rules. Machine-learning approach is to train a classifier and use the trained classifier to predict the class label. Combining the two above - the rule-based and machine learning - is the hybrid approach. In terms of question classification, the most successful approaches are Machine Learning and hybrid methods. Nguyen et al. [11] proposed a SVM based method for the same task. Firstly, the question was parsed and tokenized, parts-of-speech were tagged, data removed stop-words and was stemmed. Next they extracted a lot of features and selected features before passing the data into a support vector machine for training. For test questions they used the same pre-processing. Deep Learning approaches have proved that it is significantly beneficial for text classification, summarization, machine translation, etc. as well as for question classification [12-15]. Andreas et al. [12] proposed compose neural networks for question answering. Kim et al. [16] employed Convolutional Neural Networks (CNNs) and treated questions as general sentences to achieve remarkably strong performance in the TREC question classification task.

With regards to our research, what differentiates it from the others is the Vietnamese question answering system. This is the first research integrating two phases - diacritics restoration and question classification into Vietnamese question answering system. The diacritics restoration is the Encoder-Decoder LSTM model and Question classification uses the Bi-LSTM model. The result of the first step will be the input of question classification.

3. Methodology

Our model includes two steps: Diacritics Restoration and Question Classification. In the first step, we use Encoder-Decoder LSTM model proposed by Bui [6] and in the second step, pre-train word embeddings are used in the Bidirectional LSTM model for Vietnamese question classification. We integrate the result of the first step into question classification. An overview of our architecture is illustrated in Figure 1. We will describe more details of each layer in our model as follows:

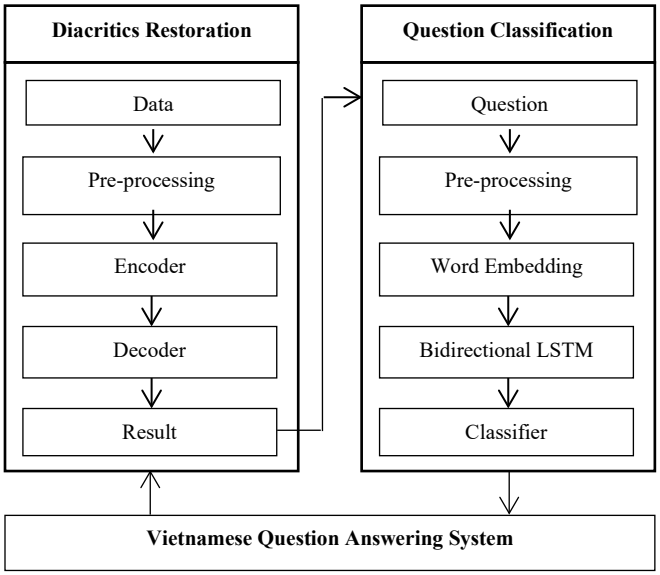


Figure 1: The proposed model

3.1. Word Embeddings

Indeed, it has been discussed for quite a long time to use vector representation for words. And recently there has been more and more interest in word embeddings - a technique which maps the words to vectors. Tomas Mikolov's Word2vec algorithm [17] is one driver for this which takes a big volume of text to create high-dimensional representations of words. By this way, it could identify the relationships between words which are not supported by external annotations. Thus, it proves to succeed in getting lots of linguistic regularities with such representation.

In this research, we use word embeddings - Continuous Bag of Words (CBOW) [17]. In the CBOW model, a variety of words illustrate the context for a defined target word which stands for a modification of neural network architecture. For example, with "trèo" (climbed) as the target word, let's use "mèo" (cat) and "trên" (on) as the words to describe the context. Figure 2 describes this model as below:

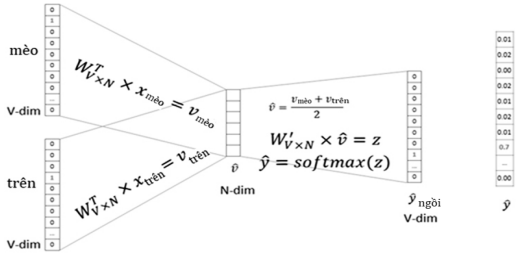


Figure 2: Continuous Bag of Words (CBOW)

3.2. LSTM

Long short-term memory (LSTM) [18] is a modification of the Recurrent neural networks (RNN). Thanks to the feedback loop in its architecture, this model has the ability to store the records of previous outputs.

Figure 3 represents a diagram of a simple LSTM cell. As befits the term deep neural networks, all single cells are combined altogether to shape a huge network. The cell unit represents the memory with five main elements: an input gate *i*, an output gate *o*, a forget gate *f*, hidden state output *h* and a recurring cell state *c*. Given a sequence of vectors (x_1 ,x_2, ..., x_n), σ is the logistic sigmoid function, the hidden state h_t of LSTM at time t is calculated as follows:

$$h_t = o_t * tanh(c_t) \quad (1)$$

$$o_t = tanh(Wx_0x_t + Wh_0h_{t-1} + Wc_0c_t + b_o) \quad (2)$$

$$c_t = f_t * c_{t-1} + i_t * tanh(Wx_cx_t + Wh_ch_{t-1} + b_c) \quad (3)$$

$$f_t = \sigma(Wx_fx_t + Wh_fh_{t-1} + Wc_fc_{t-1} + b_f) \quad (4)$$

$$i_t = \sigma(Wx_ix_t + Wh_ih_{t-1} + Wc_ic_{t-1} + b_i) \quad (5)$$

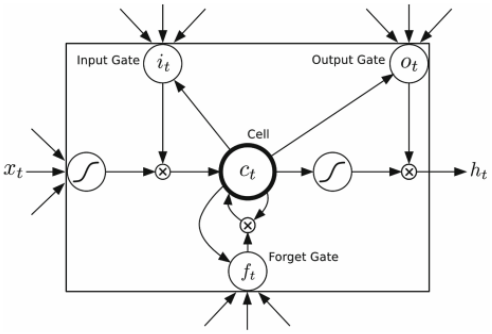


Figure 3: The Long Short Term Memory cell (Source: [19])

3.3. Bidirectional LSTM

Bidirectional LSTM [20] could be seen as a suitable model in many tasks of natural language processing. This model is modified version of LSTM by combining forward and backward LSTMs. The Bi-LSTM model is shown in Figure 4.

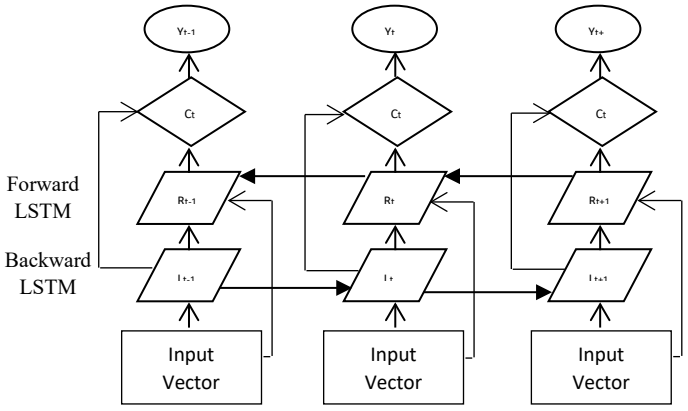


Figure 4: Bidirectional LSTM

Assuming that above equations are abbreviated to

$$h_t = \text{LSTM}(x_t\ h_{t-1}). \quad (6)$$

The forward $\overrightarrow{ht}$ and $\overleftarrow{h}$t the backward LSTM are defined as follows:

$$\overrightarrow{h_t} = \text{LSTM}(x_t, \overrightarrow{h_{t-1}}) \quad (7)$$

$$\overleftarrow{h_t} = \text{LSTM}(x_t, \overleftarrow{h_{t-1}}) \quad (8)$$

Where → denotes the forward and ← denotes backward. By taking the forward and backward LSTM as input, the Bidirectional LSTM $\overleftrightarrow{ht}$ at time *t* leads to further classification procedure:

$$\overleftrightarrow{h_t} = [\overrightarrow{h_t}, \overleftarrow{h_t}] \quad (9)$$

3.4. Diacritics Restoration

We used the Encoder-Decoder LSTM model proposed by Bui [6] for Vietnamese diacritics restoration. The Encoder-Decoder model was introduced by Kyunghyun Cho et al. [14]. This model learns to encode a variable-length sequence into a fixed-length vector representation and to decode a given fixed-length vector representation back into a variable-length sequence. This model is presented in Figure 5.

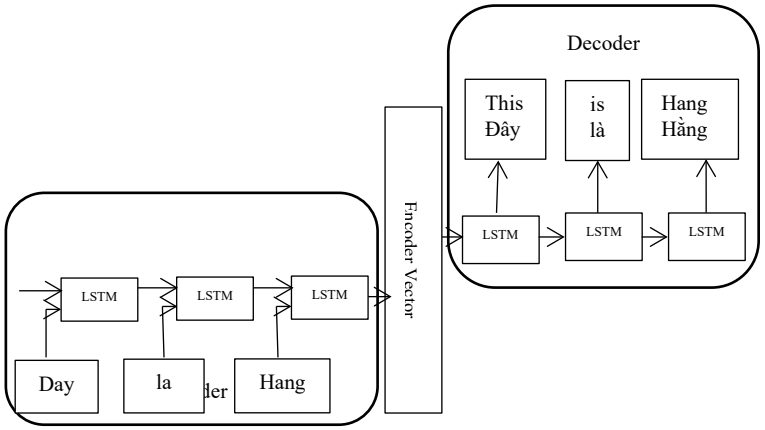


Figure 5: Encoder-Decoder LSTM model

In Figure 5, the input sequence is displayed to the network one encoded character at once. The output produced from this model is a fixed-size vector that shows the internal representation of the input sequence.

3.5. Question Classifier

The output layer is defined as follows based on the text vector studied from the Bidirectional LSTM model:

$$y = (W_d x_t + b_d) \quad (10)$$

where x_t: the text vector studied from the Bidirectional LSTM model; y: the degree of question class of the target text; W_d, b_d: the weight and bias associated with the Bidirectional LSTM model.

The Bidirectional LSTM model is trained by minimizing the mean squared error between the predictive question class and true question class. The loss function is identified as follows:

$$L(X, y) = \frac{1}{2n}\sum_{k=1}^{n}\binom{n}{k}\left\|h(x^i) - y^i\right\|^2 \quad (11)$$

$\mathbf{X} = \{x^1, x^1, x^2, \ldots, x^m\}$: a training set of text matrix

$\mathbf{y} = \{y^1, y^2, \ldots, y^m\}$: a question class ratings set of training sets

4. Experiments

We used dataset in [6] for the first step. This dataset was crawled from the official newspapers website from various categories like sports, world, business and entertainment. The size of dataset is 150 MB. The training dataset is 1500000 sentences and testing dataset is 5000 sentences. We removed all diacritical marks following the rule presented in [6].

To enable a LSTM network to process and remember more effectively, we are going to break dataset down to a lot of n-grams. Also, Vietnamese tone marks for a word can be determined from the small surrounding words in most of the cases; therefore something like a 5-gram or 7-gram model will rather fit for our purpose to some extent. So we have 11 million 5-gram with length varied mostly from 15-25 characters. In addition, as mentioned above, the easiest way to gather a lot of training data to add diacritical marks to Vietnamese text is to get a diacritical text then remove the diacritical marks.

As a machine learning framework, Keras with Tensorflow as a backend has been used. We used same parameters of [6] with 3 LSTM layers, each layer has 256 neurons. We optimized by Adam Optimizer with learning rate 0.002. We used Accuracy score - Accuracy of Diacritics Restoration (ADR) to evaluate the first step as below equation:

$$ADR = \frac{Number\ of\ corrected\ words}{Number\ of\ words} \quad (12)$$

We did experiements with the range of epochs, after about 300 more epochs, or 16 hours we got 97% accuracy and reduced loss to 0.07. Figure 6 shows the Loss of Diacritics Restoration over epochs and Accuracy of Diacritics Restoration per epochs are shown in Figure 7.

Since the used data was extracted from the newspaper websites, it might consist of unpopular words and a couple of errors. Therefore, the diacritic restoration in those cases usually will not achieve highly accurate results.

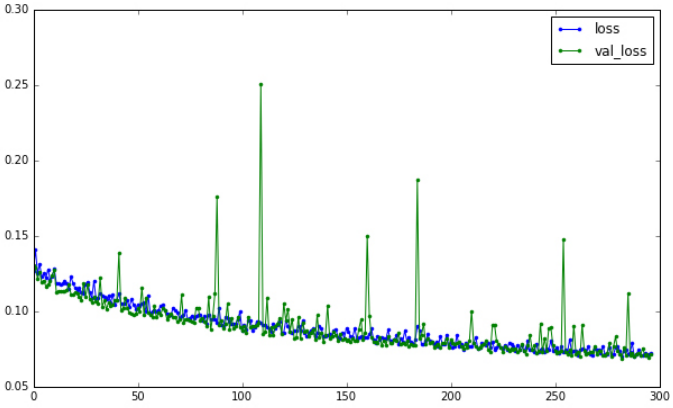


Figure 6: The loss of Diacritics Restoration over epochs

For the second task, we collected question classification dataset from Binh Duong Department of Information and Communications. This data includes 4 areas with 42 types of provincial administrative procedures [21]. The dataset has 560 questions labelled as five primary categories and 270 sub-categories. The main categories and their corresponding number of subcategories are listed in the Table 1. Table 2 shows the details our dataset for training and testing in Question Classification with ratio 8:2.

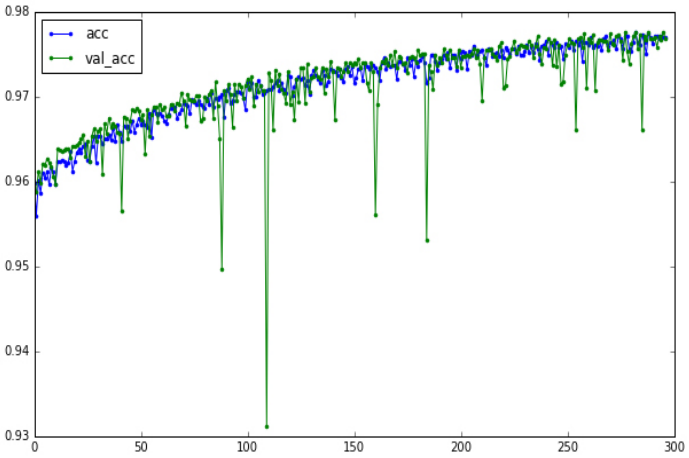


Figure 7: Accuracy of Diacritics Restoration task per epochs

Table 1: Describe of dataset of Question Classification

Categories	Number of Sub-Categories
Journalism	35
Postal	42
Broadcasting	84
Publishing	98
General Information	11

Table 2: Dataset of Question Classification

Name of Dataset	Number of Sentence
The set of questions	570
Sub- Categories	270
Train	456
Test	114

For pre-processing dataset, we used Pyvi (0.0.0.9 - Tran Viet Trung 2016) to tokenize Vietnamese, pre-train Vietnamese word embeddings by streetcodevn (Hung Le 2018). For our training models, we used Tensorflow and Keras libraries. The parameters of our question classification Bi-LSTM model are: Batch size: 500, number of hidden nodes: 128, Drop out: 0.2, epochs: 300, Sigmoid Activate function, Adam Optimization function and Binary cross entropy function. We evaluated performance of question classification by Accuracy of Question Classification (AQC) following below equation:

$$AQC = \frac{\#of\ correct\ predict\ questions}{\#of\ predict\ questions} \quad (13)$$

To evaluate our proposed model, we compared our results with the result of LSTM model separately as illustrated in Table 3. This

result reveals that our proposed model – the Bi-LSTM deep learning approach get the best results. From the result is shown in Table 3, it is obvious that the Bi-LSTM beats the LSTM model; therefore it could be said that the Bi-LSTM is the best-performing score comparing with the LSTM.

Table 3: The results of question classification

Model	AQC
LSTM	92%
Bi-LSTM	**95%**

We made an application by integrating the automated question answering systems into Binh Duong Department of Information and Communications Support System. The name of this application is ICTbot. The ICTbot is used to support Binh Duong Department of Information and Communications in General Information and four management categories: Journalism, Postal, Broadcasting and Publishing. The ICTbot allows users to input new questions and the result is the answers that have the highest accuracy of question classification. Figure 8 shows the ICTbot application.

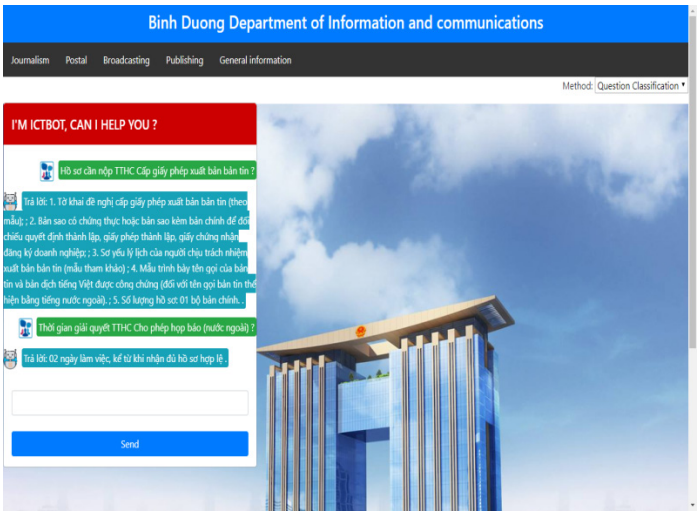


Figure 8: The ICTbot of Binh Duong Department of Information and Communications Support System.

Table 4: The results of question classification in different percentage of Accuracy of Diacritics Restoration

Percentage of Accuracy of Diacritics Restoration	AQC
>= 90%	95%
<= 75 and <90%	70% - 94%
<= 50 and <75%	48% - 69%
<= 25 and <50%	20% -47%
<25%	0-19%

To evaluate the differences in integrating diacritics restoration into question classification, we applied diacritics restoration by percentage into question classification. We counted the number of restoration words in test question dataset and integrated diacritics restoration in test question dataset by percentage of Accuracy of Diacritics Restoration. When percentage of diacritics restoration is equal or larger than 75%, question classification worked well, however when the percentage was so small question classification didn't work well or made a false label of the predict question. The result also depends on the length of question test. Table 4 reveals the result in different percentage of Accuracy of Diacritics Restoration. As a result, it could be seen that integrating diacritics restoration into question classification improves the accuracy of the system.

5. Conclusion

In this paper, we experimented our model by integrating diacritics restoration and question classification into Vietnamese question answering system. The result of the first step will be the input of question classification. It could be said that this is the first research integrating two phases into Vietnamese question answering system. We used the Encoder-Decoder LSTM model for Vietnamese diacritics restoration and the Bidirectional LSTM model for Vietnamese question classification. Our experiments proved that integrating diacritics restoration improves question classification. We also built a useful application based on Question classification – ICTbot and integrated in Binh Duong Department of Information and Communications Support System. In the future, we are looking to apply other deep learning approach and explore more features to improve the results.

Conflict of Interest

I declare no conflict of interest.

References

[1] G. D. Pauw, P. W. Wagacha, and G.-M. de Schryver, "Automatic diacritic restoration for resource-scarce languages," in Proceedings of the Text, Speech and Dialogue, 10th International Conference, TSD, Pilsen,Czech Republic, pp. 170–179, 2007. https://doi.org/10.1007/978-3-540-74628-7_24

[2] M. Simard and A. Deslauriers, "Real-time automatic insertion of accents in French text," Natural Language Engineering , vol. 7, Issue 2, pp. 143–165, 2001. https://doi.org/10.1017/S1351324901002650

[3] Mihalcea, R, "Diacritics restoration: Learning from letters versus learning from words", in Proceedings of CICLing, pp. 339-348. Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing, 2002. https://doi.org/10.1007/3-540-45715-1_35

[4] Nikola Santic, Jan Snajder, Bojana Dalbelo Basic, "Automatic diacritics restoration in Croatian texts", The Future of Information Sciences, Digital Resources and Knowledge Sharing, pp. 126–130, Springer, 2009.

[5] R. Nelken and S. M. Shieber, "Arabic diacritization using weighted finite-state transducers", in Proceedings of the ACL Workshop on Computational Approaches to Semitic Languages, ser. Semitic '05, Stroudsburg, PA, USA, pp. 79–86, 2005. https://doi.org/10.3115/1621787.1621802

[6] Bui Thanh Hung, "Vietnamese Diacritics Restoration Using Deep Learning Approach", in Proceedings of the 10th International Conference on Knowledge and Systems Engineering – KSE, 2018. https://doi.org/10.1109/KSE.2018.8573427

[7] L.N. Pham, T.V. Hong, V.V. Nguyen, "Vietnamese Text Accent Restoration with Statistical Machine Translation", 27th Pacific Asia Conference on Language, Information, and Computation pp.423 — 429, 2013. https://www.aclweb.org/anthology/Y13-1044

[8] N.M.Trung, N.Q.Nhan, N.H.Phuong, "Vietnamese diacritics restoration as sequential tagging", Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), IEEE RIVF International Conference, 2012. https://doi.org/10.1109/KSE.2018.8573427

[9] T.A. Luu, K. Yamamoto, "A Pointwise approach for Vietnamese automatic diacrics restoration", Proceedings of the International Conference on Asian Language Processing-IALP, 2012. https://doi.org/10.1109/IALP.2012.18

[10] Amit Mishra and Sanjay Kumar Jain, "A survey on question answering systems with classification. Journal of King Saud University - Computer and Information Sciences, 28(3), pp 345 – 361, 2016. https://doi.org/10.1016/j.jksuci.2014.10.007

[11] Nguyen Van-Tu and Le Anh-Cuong, "Improving question classification by feature extraction and selection", Indian Journal of Science and Technology, 9(17), 2016. https://doi.org/10.17485/ijst/2016/v9i17/93160

[12] Andreas, J., Rohrbach, M., Darrell, T., and Klein, "Learning to Compose Neural Networks for Question Answering", Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016. https://doi.org/10.18653/v1/N16-1181

[13] Graves, A., "Supervised Sequence Labelling with Recurrent Neural Networks", Studies in Computational Intelligence, vol. 385, pp. 5-13, Springer, 2012.

[14] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, Yoshua Bengio, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation", Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1724–1734, 2014. https://doi.org/10.3115/v1/d14-1179

[15] Tom Young, Devamanyu Hazarika, Soujanya Poria, Erik Cambria, "Recent Trends in Deep Learning Based Natural Language Processing", IEEE Computational Intelligence Magazin, 2018. https://doi.org/10.1109/MCI.2018.2840738

[16] Yoon Kim, "Convolutional neural networks for sentence classification", Conference on Empirical Methods in Natural Language Processing, pages 1746–1751, 2014. https://doi.org/10.3115/v1/D14-1181

[17] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean, "Efficient estimation of word representations in vector space", in Proceedings of International Conference on Learning Representations (ICLR-13): Workshop Track, 2013. https://arxiv.org/abs/1301.3781

[18] Hochreiter S., Schmidhuber J., "Long Short Term Memory", Neural Computation 9(8), pp. 1735–1780, 1997. https://doi.org/10.1162/neco.1997.9.8.1735

[19] Hong J, Fang M, "Sentiment analysis with deeply learned distributed representations of variable length texts", Technical report, Stanford University, CS224d: Deep Learning for Natural Language Processing, 2015. https://cs224d.stanford.edu/reports/HongJames.pdf

[20] Wang P., Qian Y., Soong F. K., He L., Zhao H., "Part-of-Speech Tagging with Bidirectional Long Short-Term Memory Recurrent Neural Network", 2015. https://arxiv.org/abs/1510.06168

[21] Decision No. 1402/QD-UBND dated May 29, 2018 on the announcement of administrative procedures under the jurisdiction of the Department of Information and Communications/District People's Committees of Binh Duong province.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 213-223 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Improvement opportunities of a Simulation/Expert System Approach for Manufacturing System Sizing: A review and proposal

Lajmi Imen[1,*], Masmoudi Wassim [2], Elleuch Mounir[2], Chtourou Hedi [3]

[1]*University of Sfax, Research Unit of Industrial Management and Logistics Quality (LOGIQ) Higher Institute of Industrial Management Sfax, Tunisia*

[2]*University of Sfax, Laboratoire des Systèmes Electro-Mécaniques (LASEM), Ecole Nationale d'Ingénieurs de Sfax, Sfax, Tunisia*

[3]*University of sfax, Institut Préparatoire aux Etudes d'Ingénieurs de Sfax, département de Technologie, Sfax, Tunisia*

ARTICLE INFO

Article history:
Received: 03 June, 2019
Accepted: 13 September, 2019
Online: 08 October, 2019

Keywords:
Manufacturing system sizing
Handling Resources
Production resources
Simulation
Expert System
Performance

ABSTRACT

Manufacturing System (MS) sizing is a crucial task to complete in order to obtain the desired MS performance and efficiency. It involves selecting the required number of resources from each used type in a given planning horizon. In fact, different approaches coupling simulation/optimization tools have been developed to solve this issue and evaluate the MS performance. One of these approaches is the Simulation Expert System Approach (SESA). Unfortunately, the application domain of this approach is limited in sizing only the production resources (machines and labor) but neglects the material handling system (MHS) components. Moreover, omitting the transferring problem is not viable in the real world due to its importance in each shop floor. Thus, the aim of this paper is to describe the evolution of SESA, then, to check if the simulation optimization tools used in SESA are still relevant. This paper also investigates the importance of incorporating MHS in this approach and finally proposes some improvement opportunities for SESA including the tackling of the MHS fleet sizing problem. In fact, the wide literature review performed in this research indicates that SESA is still a pertinent approach but it must be improved. Therefore, it is expected that SESA improvement opportunities proposed in this work will greatly assist industrialists in enhancing the overall MS performance, providing a significant productivity increase and a minimization of the total production costs.

1. Introduction

This paper is an extension of a work presented in the 6th IEEE International Conference on Advanced Logistics and Transport (ICALT) [1].

Due to intense competitivity in the industrial sector and to demand fluctuation, companies need to ensure better performance of their Manufacturing Systems (MS). An appropriately selected number of resources can improve the overall system performance. On the other hand, too much chosen resources means needless investment, whereas an insufficient number of chosen resources can make the system unable to produce the required quantities in the desired due date, which constitutes a financial loss (ie. penalties, loss of customers, etc.). Therefore, MS sizing is an important issue that must be considered to obtain the desired system performance. In fact, this issue has attracted the attention of several researchers. Hence, many types of approaches have been used to solve machines, labor and MHS fleet sizing problems. The Simulation/Expert System Approach (SESA), is one of these approaches used to solve MS sizing. It was proposed by [2] and enhanced by other researchers either by considering other resources or by altering the used tools. However, SESA still needs improvements to be applicable in real cases. However, one of its shortcomings is the lack of consideration for the material handling system (MHS). In fact, appropriate MHS choice and sizing is crucial for the MS design task [3]. Its main role is transferring parts between the various components of a MS in the most economical way, to ensure an efficient material flow. The interest of configuring and sizing the MHS is continually increasing in

*Corresponding Author Imen Lajmi, Sfax 3062, Tunisia, +21652486285, Email: lajmi.mezghani.imen@gmail.com

manufacturing facilities. Therefore, this research proposes some opportunities to improve SESA framework considering the MHS fleet sizing problem to make this approach more pertinent and applicable in real-world cases.

The remainder of this paper is as follows: Section 2 presents a brief literature review on the MS sizing problems. The originality of SESA and its evolution are presented in Section 3. Section 4 presents a survey of recent research studies to show the promising potential of SESA tools. Before concluding, section 5 suggests an improvement proposal for SESA.

2. MS Sizing problem

MS sizing problem is defined as the selection of the required number of resources from each type to be used in a manufacturing process for a given horizon.

2.1 Resources studied in literature

Several research studies dealt with the MS sizing problem, more precisely with the main manufacturing resources used in MS, such as (1) machines, (2) labor and (3) material handling system (MHS).

2.2 Machine selection problem

Machines represent a fundamental industrial equipment that allocates a part, for a period, to generate changes by performing certain tasks (generating the Added Value). Finding the required number of machines is an important task in the MS to achieve the target performances. Among the first research studies discussing the machine selection problem, we can cite [4] who developed neural networks to minimize the number of machines in each work center in a flexible manufacturing system. In addition, [5] proposed a mathematical formulation and a heuristic algorithm to find out the required number of types of machines in a group technology system that reduce the manufacturing cost. More recently, [6] developed a new modelling approach combining simulation/optimization methods to find the minimum number of machines required in each MS type. Besides,[7] used an artificial neural network approach coupled with a simulation tool to estimate the minimum number and the best scheduling of machines that enhance the mean tardiness and the mean flow time. In fact, [8] used two analytical methods to firstly, determine the optimal number of machines and the quantities produced in each period and secondly, determine the production plan and the periodic preventive maintenance scheduling to reduce the failures and minimize the total costs. Moreover, [9] used Discrete-Event Simulation (DES) and OptQuest to find out the optimal machine number. Two objective functions were used: maximizing the machine utilization rate and minimizing the waiting time in the queue.

The continuous dealing with the machine selection problem proves the importance of this issue in the MS improvement.

2.3 Labor selection problem

The human factor plays a crucial role in the creation, the responsiveness, the flexibility and the performance of industrial processes. The determination of the optimal number and quality of labor to be allocated is necessary in each MS. In fact, this issue has been the object of numerous research studies. For instance, [10] used Fuzzy AHP, TOPSIS and Simulation tools to determine the optimal number and assignment of labor in the system. Many performance measures have been used to achieve some objectives such as the average lead-time, the average labor utilization and the average waiting time. [11] used simulation and genetic algorithm to determine the optimum number and allocation of labor and the measure of the efficiency of their operation in an assembly shop. His aim is to maximize the throughput system. Similarly, [12]developed a simulation /optimization approach based on the genetic algorithm (NSGAII) to determine the required number of labor. [13] used CPLEX optimization software and DES (ARENA software) to determine the required number of operators in automobile chassis manufacturer while considering the labor fatigue in their attempt to satisfy the customer demand.

2.4 MHS fleet sizing

MHSs are used to transport goods and materials between workstations of a manufacturing system. Therefore, the number of vehicles greatly influences the performance of the MS. MHS are usually expensive, thus determining the appropriate number of vehicles is very a very important task. Hence, due to the importance of this problem, many recent research works dealt with this problem using different approaches. For instance, [3] developed an analytical model to estimate the minimum number of automated vehicles required by a flexible MS. The results of the regression analysis were compared with simulation results to show that the regression technique is a promising approach to resolve the fleet sizing problem. [14] used a simulation tool to determine the optimal vehicle number and material flow rate in a semiconductor manufacturing systems. [15] applied a queuing model to determine the required number of automated vehicles in a flexible MS containing multiple pickup and delivery points. The objective was to minimize the waiting times. As for, [16] applied the queuing theory to specify the required number of MHSs in a MS in order to maximize their utilization rate and minimize the investment costs. Besides, [17] used the bees algorithm to estimate the optimal number of MHS operated at the MS to maximize the profit. This algorithm is inspired by the foraging behavior of honeybees. On the other hand, [18] used multi-class Closed Queuing Networks (CQN) based on a linear programming to model the movement of MHS in a MS and determine the minimum MHS number. More recently, [19] has developed two analytical methods to determine the AGV fleet size in different layouts of flexible MS. In fact, different factors, such as the processing time, job sequence, job mix, loading/unloading stations and the number of work centers, are included in the model.

2.5 Synthesis

Table 1 presents a literature survey on the MS sizing problem the detailed analysis of which shows that the number of studies about the production resources selection area has grown considerably in recent years. In fact, the growing interest in this

subject shows the importance of the resolution of the resource selection problem in improving MS performance. However, one of the gaps in these studies is the absence of a correlation between the three types problems related to the selection of machines, of labor (operators) and of MHS components, which affect one another.

Table 1. Literature review on MS sizing

Types of resources	Research studies
Machine	[4-9]
Employees	[10-13]
MHS	[14-19]
Machine and employees	[20, 21]

Recently, some studies have combined machine and labor selection problems but without considering the MHS fleet sizing problem. For example, [20] used a simulation tool to size an effective automobile manufacturing process. The aim is to enhance the system throughput by determining the required number of repair stations, labor, machines and the best plan configuration. More recently, [21] has sized a functional MS to determine the required number of machines and employees that enhance the overall system performances for the purpose of respecting the Due Date (DD).

2.6 Used methods

Due to the variety of resources used in the production systems and the influence of one over the other, the search for the optimal size becomes more and more difficult. To overcome this difficulty, many methods are employed in the literature to tackle the sizing issue, which are classified into two main categories:

- Analytical methods;
- Simulation-based methods.

The first category can be classified into two main types: exact and approximate methods.

In fact, the exact methods like Branch-and-Bound, linear programming and others provide an optimal solution, which is suitable only for small sized problems. Therefore, the major weaknesses of these methods are their static and deterministic character, their mis-consideration of the dynamic and stochastic aspects of the system and their inconsistency with the MS sizing problem. Since the number of feasible solutions rises exponentially when changing the total number of resources, it is impossible to use complete enumeration for large-sized problems. As a consequence, heuristic and meta-heuristic methods are widely approved by the researchers to efficiency deal with the MS sizing problems.

Approximate methods, which are heuristic and artificial intelligence methods, as the expert system, artificial neural networks, genetic algorithm, Tabu search, particle swarm optimization, simulated annealing and others do not always guarantee the optimal solution.

On the other hand, the simulation-based methods, which fall into the second category, are widely used to solve MS sizing problems. In fact, simulation can provide a detailed analysis of the system performance and help managers to take the right decision.

Moreover, some studies developed approaches combining simulation and analytical methods to find out satisfying results (not optimal ones). Due to the effectiveness of this combination, [21] used an expert system simulation approach called "SESA" to look for the optimal number of two types of resources simultaneously: machines and labor. In fact, this work considers the stochastic and dynamic aspects. However, the major gap in this work is the neglect of transfer problem when sizing MS.

In the next section, the evolution of SESA and its originality are presented.

3. SESA Evolution

3.1. Originality of SESA

SESA is a prescriptive simulated approach developed by [2] to solve the MS sizing problem and more precisely, to identify the optimal number of machines in a simple static and deterministic MS. The framework of SESA is presented in Figure 1. The two principal tools used in SESA method are the simulation and the expert system, which work in a closed loop and exchange data and results.

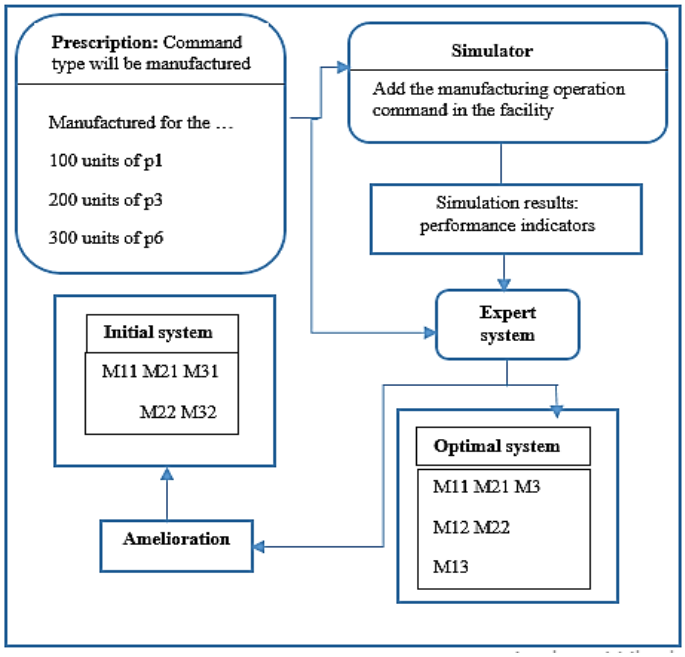


Figure 1. The framework of SESA [2]

The data of the studied MS and the typical order to be manufactured are the required inputs for the simulation step in order to establish the simulation procedure according to which the command type will be executed. This first step provides performance indicators such as the total operation time. For each scenario, the simulation model is run and the system performance measures are recorded. Then, the second step, it involves the expert system, which has as inputs the performance measures (provided

by simulation) and the performance limits. The main role of this step is to analyze the results provided by the simulator and establish one of these two cases:

- The system performance, which is optimized according to the actual context, or;
- The system performance, which needs to be improved according to the actual context. In fact, the expert system offers a problem diagnosis, according to the MS resources, in order to check if there is a lack or an excess of resources. Then, the ES provides a set of recommendations to adjust the resource number, which improves the system performance.

The simulator used in this original version is a deterministic tool named "SICOP". It is a simple program developed by "C language", which provides statistical results (performance measures). The expert system is "SEORAM", which is intended to minimize the Mean Flow Time (MFT). The results of this study indicate that combining the simulation and optimization tool is quite competitive, in terms of precision, when compared to the simulation itself. This approach helps to avoid the "try-error" aspect. Enhancement of the system performance proves the efficiency of this approach. Accordingly, other researches were encouraged to extend the application of this approach to be more applicable in real MS.
The weaknesses of this study are as follows:

- Misconsideration of the stochastic and dynamic aspects;
- Used tools and the total indices of performance chosen in this initiated approach are not used any more in research since they do not reflect any more the current tendencies of the industrialists;
- Misconsideration of the labor and MHS selection problems.

3.2. SESA improvement

Improvement of the Expert System used tool "ESMRS"

In the follow-up research work, [22] developed an improved version of SESA and applied it for Job Shop machine selection problem. The studied MS grouped the same machine type on departments and supposed that the requirement of labor and material handling system are neglected. The main focus was on the development of a new ES prototype named expert system for manufacturing resource sizing (ESMRS developed by Kappa. PC) (see Figure 2). A set of generic diagnosis production rules is developed. The main objectives were to improve the structure of research mechanism, recommend resource modification according to performance indicators provided by the simulation and overcome the "try-error" aspect. The results provide significant insights into the effective application of the enhanced ES.

Improvement of the simulation used tool

After that, [23]enhanced the framework of SESA, previously described, by using the simulation software "ARENA". This tool considered the stochastic and dynamic aspects, governs Manufacturing Orders launching (MO) as well as machine operating parameters. Thus, all entities of the same product type enter the MS according to a stochastic inter-arrival time (random distribution). Then, the entities stay on the machine queue until the machine becomes available. After that, the entity will be treated in a randomly chosen processing time (see Figure 3). The same optimization loop presented previously is used in [23] but the differences are:

- Use of a stochastic simulator tool;
- Use of more relevant performance indicators and;
- Considering the stochastic and dynamic aspect.

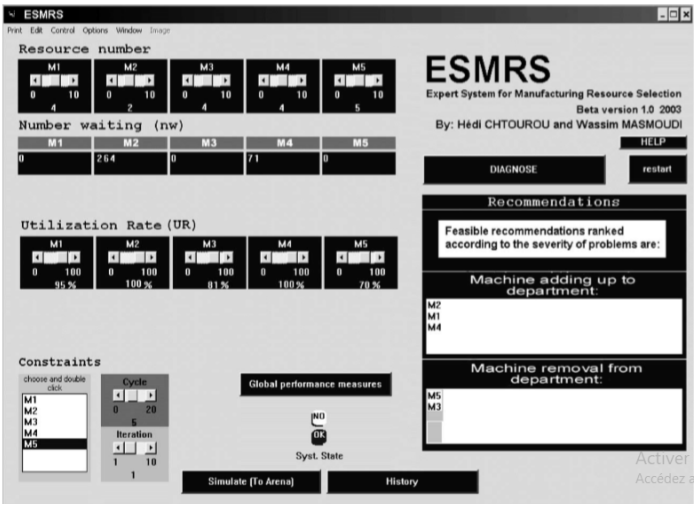


Figure 2. Enhanced ES interface [22]

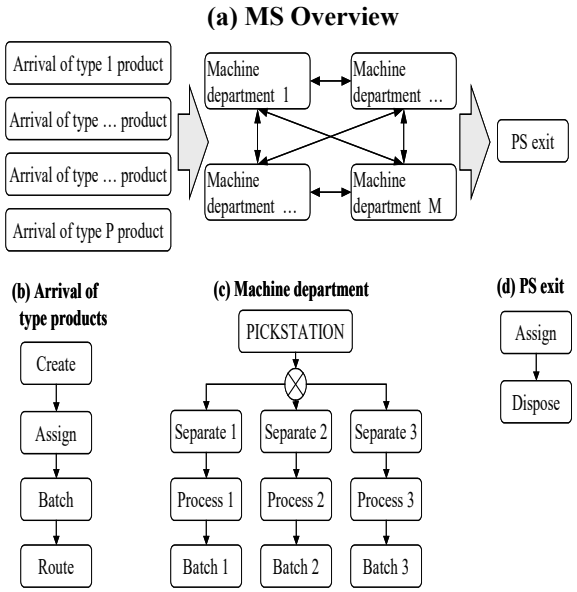


Figure 3. MS "ARENA" simulation model[23]

Despite improvements provided by [22, 23] to enhance the efficiency of SESA tools, this approach still limited in its application domain. It resolve only the machine selection problem.

3.3. Enlargement of SESA application domain

In an earlier work, [21] extended the application domain of SESA to select the optimal number of both machines and labor in a functional manufacturing system. This extension has already involved changes both in simulation model (see Figure 4) and ES interface (see Figure 5).

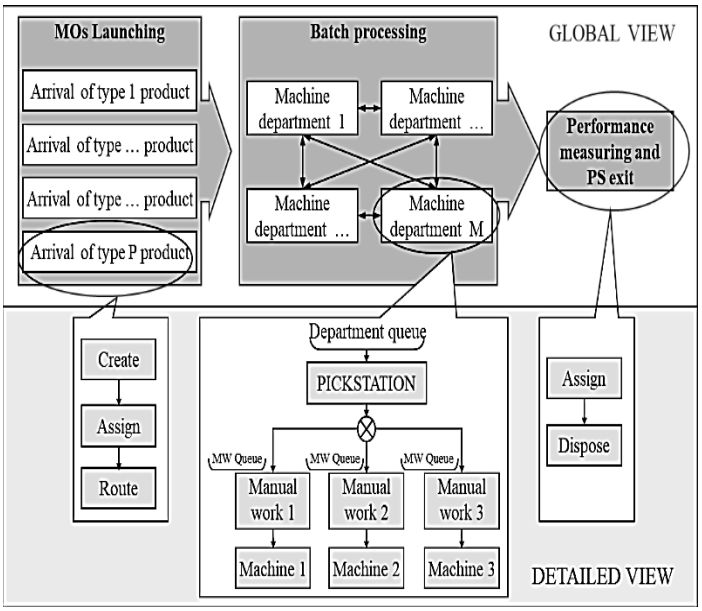


Figure 4. MS "ARENA" simulation model [21]

To build the simulation model, two inputs are required:
The MS data which represents the number of work center, Processing Time (PT) and Setup Time (ST) and;
The demand patterns, which represent the inter-arrival times, product type, load size and sequence.

Different performance indicators have been used to enhance the applicability of SESA. These indicators are: the rate of resource utilization, the average waiting time and waiting number of batches in resource queues, throughput, mean tardiness and mean earliness.

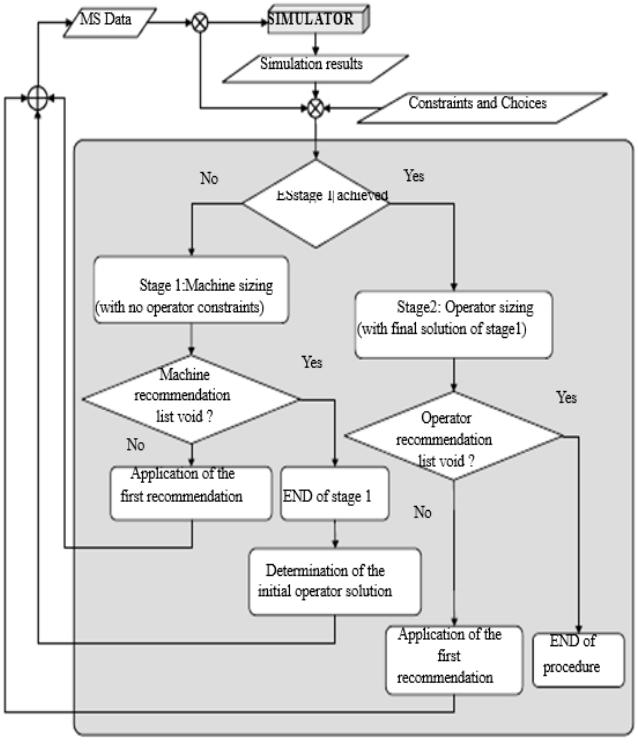


Figure 5. Enhanced ES interface [21]

The main target is to respect the DD in order to minimize the delay penalties and the storage costs. The DD indicator is more adequate for functional manufacturing context than MFT. The first objective is to minimize the Mean Tardiness (MT) and then minimize the Mean Earliness (ME), which is considered as the second objective. Besides, in order to avoid superfluous investment costs, all the resources should be fully utilized. The experiments with different levels of DD closeness indicate the important benefits of the approach improvements through the new performance measures and the enhanced ES interface.

The enhanced ES interface is divided into two main stages. The first stage is the machine selection problem without labor constraints and the second is the labor selection problem.

In the first stage, the ES starts with the study of the MS machine situation (without labor constraints). Thus, this stage supposes that the number of employees is equal to that of machines (assigning a worker to each machine). If the simulated system performance needs to be improved, the ES will recommend modifications regarding machine number. These changes overcome the problem and enhance the system performance. Thus, a new cycle is run with the modified machine number until the ES can not suggest other modifications. This situation presents the end of this stage. In the second stage, the ES resolves the labor selection problem with a machine balanced MS (according to the first stage results). The same optimization cycles will be run until they reach a non-improvable MS.

The confines of SESA developed by [21] is the neglect of the transfer network and MHS constraints when sizing MS component resources. In fact, MHS is significant in MS that has an effect on the system performance [24]. It can account for 30 –75% of the total cost, and efficient material handling can be primarily responsible for reducing a plant operating cost by 15–30% [25]. Therefore, MHS is very important in reducing the total manufacturing costs and increasing profits. Consequently, it is necessary to extend the application domain of SESA by incorporating material handling system constraints when sizing production resources. However, it is necessary to demonstrate the relevance of SESA tools. Therefore, the next section will provide a detailed survey of recent research studies using simulation and/or expert system for solving resource selection problem.

4. Is SESA a promising approach?

4.1. Is simulation always promising?

Simulation has been developing for a long time to became an essential tool for many disciplines. Its application fields may go from industry [26] healthcare [27, 28], urban design [29] military [30], among others. It has been widely proven by the scientific community and practitioners as a flexible technique by allowing modelling, testing and analyzing complex dynamic systems, namely for models where structure and behavior change over time. The simulation model is a representation of the real system. It may reduce the risks related to the implementation of new system design, production cost, production time, etc. It is used to analyze

and process huge quantities of data. As shown in Figure 6, simulation is a solid approach with many years of application and is a rapidly evolving research field [31].

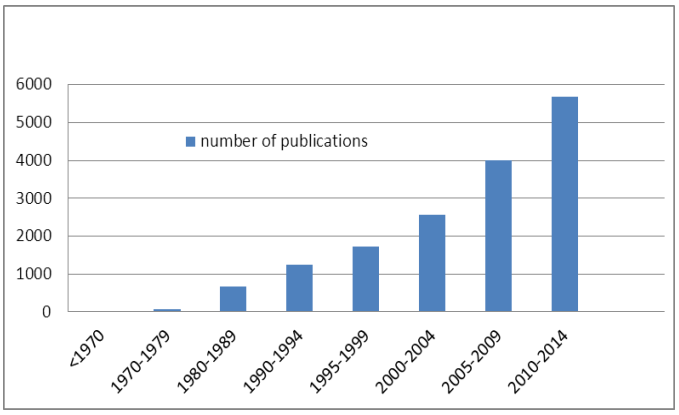


Figure 6. Evolution of simulation application

Discrete-event-simulation (DES) is mostly used to solve MS problems. Numerous benefits to use DES defined by [32]include: simulate a real system in a short time, study the interaction between different components, reduce cost and risk, monitor and control the system and others.

This approach has tackled different problems. A brief summary of DES applications to solve manufacturing systems is presented in Table 2. [33] revealed abundant references in more details about the simulation utilisation on other MS problems.

Table 2. Publications related to simulation approach

Authors	Years	Simulation application
[34]	2008	System and facility design
[35, 36]	2015, 2017	Planning and scheduling
[37]	2016	Production planning and control
[38]	2019	Supply chain design
[39]	2016	Inventory management
[40, 41]	2019, 2018	Maintenance
[42]	2016	Controlling and analyzing energy and power consumption
[43]	2018	Resource allocation

In all the above discussed studies, DES models have been proven as powerful decision support tools, which enables practitioners to accurately test, model, analyze and plan complex systems. Moreover, it is an efficient tool to search for optimal solution and evaluate the obtained results. Different DES tools have been used in literature.in this context, [44] compared these tools to identify the most popular and used ones. In fact, the ARENA commercial software has been the most popular and used tool for modeling and analyzing various complex MSs.

Nevertheless, the simulation tool may not be the first preferred in solving real-life system due to the "trial-and-error" nature of the development process and its lengthy computational requirements. This issue is particularly remarked when the solution space is mostly large and the time available to make decision and analysis is too limited. Thus, it is necessary to add an analytical method to the simulation tool to do an optimal decision-making. Therefore, the simulation tool has been coupled with an expert system approach.

4.2. Is the Expert system still promising?

The ES is an Artificial intelligence (AI) software application that attempts to simulate the knowledge and experience of human experts in a specific domain and make such expertise accessible on demand to the program user. This application was introduced by [45]for the first time in the 1980s. A growing interest in the use of ES has been seen for solving optimization problems in many domains such as Healthcare [46, 47], supply chain [48],transport [49], risk management [50] and others.

According to [51], the expert system includes three principal parts (Figure 7):

- User's interface that enables the user to ask questions and provide advice corresponding to it;
- Knowledge base that consists of facts and rules, which are created from information which are provided to it;
- Inference engine that helps to match the user's query with knowledge base and provides the result.

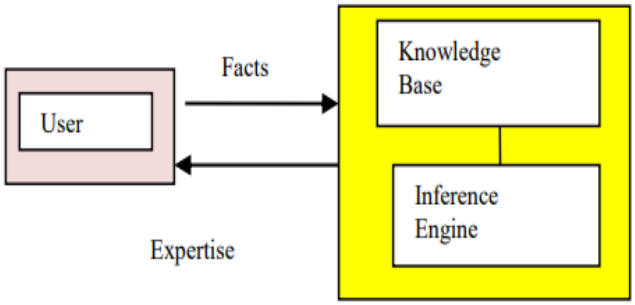


Figure 7. An Expert system environment

Table 3. Recent research studies used ES

Authors	Years	ES applications
[52]	2016	Evaluation of the performance of sustainable manufacturing
[53]	2016	Material handling equipment selection
[54]	2019	
[55]	2019	Layout designs
[56]	2018	Optimization of Design and Manufacturing Process
[57]	2019	Shift work
[58]	2015	Machine tool selection
[59]	2015	Material Selection
[53]	2016	
[60]	2017	
[61]	2018	Risk management

The main advantages of ES consist in reducing human error, always available, providing expertise at a minimum cost, which is used in any risky environment to solve complex decision problems

in different specific fields, obtain flexible and practical solutions, and respond at great speed, etc.

The capability of ESs to acquire performance similar to the human ones makes current researchers used it to solve various types of problem in MS (see Table 3).

4.3. Is coupling Simulation/Expert System promising?

Simulator systems generate large and a high variety of data that make decision making so difficult. Therefore, novel approaches combined simulation and optimization have been developed in literature to eliminate this gap. In fact, simulation based optimization approaches have received considerable attention from many recent researchers to solve MS problems. [62] proposed an improvement framework using genetic algorithm and DES to investigate the buffer size and job sequencing optimization problems. [63] combined the DES models and artificial neural network models of the production system. This combination was realized through an ES in order to evaluate the stability of a production system and implement production plans. [64] developed a practical approach using bee colony to improve throughput while minimizing waste energy and environmental impact. [65] proposed a novel methodology using ant colony and DES to select better facility allocations that minimize fixed and transportation costs.

Recent research studies applied the DES with ES in different application including industrial systems [66], food process [67], supply chain management [68], healthcare systems [69], risk management [70]and others.

The main advantage of this combination is that the DES can integrate compound system dynamics and uncertainties. Then, ES helps in efficiently getting optimal parameter values. The methodology is that the ES accesses the output of a simulation cycle and, depending on the user's objective, selects and retrieves the pertinent data and provides a mechanism for exception reporting. A greater degree of integration would be for the ES to explain the recommendations. Thus, the incorporation of DES with ES improves the decision-making process.

The research studies, cited in the previous three sub-sections, have been conducted to assess the trends of SESA approach and its capability to solve highly complex systems with stochastic variables. We can conclude that SESA is still promising but can be improved.

4.4. Limits of SESA

The application domain of SESA is limited in finding the optimal number of machines and employees without taking into account the MHS, despite its importance. A MHS can be defined as a vehicle that moves work-parts at short distances and usually inside the building. [71] affirms that the MHS represents a major portion of the total manufactured product cost. An efficient MHS can have a main impact on the operating cost, which reduces and increases the total throughput and department utilization. Moreover, a well selection of the required number of MHS is very important to significantly improve the efficiency of the MS. In fact, if the chosen number of MHS is less than the required number, the work-parts will saturate the storage areas between departments and their waiting time increases. Moreover, if the chosen number is higher than the required one, the MHS waiting time increases. This will cause an accumulation of vehicles, a saturation of the traffic lines and a system blockage. Besides, the MHS purchase cost is very high, so an accurate determination of their optimum number is required.

Consequently, considering MHS when selecting machines and labor, it is very important to significantly improve the efficiency of MS.

5. SESA improvement proposal

From previous sections, it is clear that SESA is always promising seen simulation and ES benefits. Nevertheless, it is a priority to take into consideration the MHS. Our contribution in this section is to highlight the opportunities of SESA improvements.

5.1. Proposal to enlarge the application domain

To enlarge the application domain of SESA, we will propose some improvement opportunities and develop a new strategy of SESA by considering the MHS fleet sizing problem (see Figure 8).

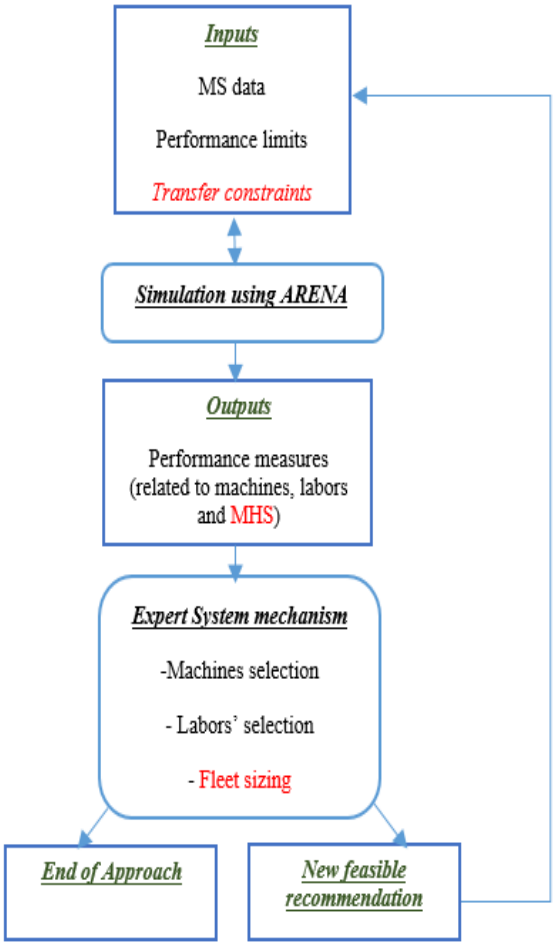


Figure 8. The new SESA framework

In the beginning, different characteristics related to the MHS will be integrated in simulation inputs. Then, new performance indicators related to the MHS will be considered to analyze system performance (see section 5.4). Finally, a new stage related to the MHS fleet sizing will be added to extend the expert system mechanism.

5.2. Improvement proposal of simulation model

To improve the simulation model developed by [21] and make the first investigation of taking into account the MHS when sizing MS, it is necessary to add different data related to the MHS within the MS data, which represent the simulation input. A representation of the improved model is presented in Figure 9.

The MHS data requirements can be:

- Type of material handling;
- Direction of movement (orientation);
- Transport distance;
- Load size;
- Loading and unloading time;
- Speed of material handling;
- Dispatching rule.

The components of this proposed model are as follows:

- Materials warehouse;
- Five departments (each one group the machinery with the same process capability) ;
- Parking (group the MHS);
- Finished Pieces (FP) warehouse.

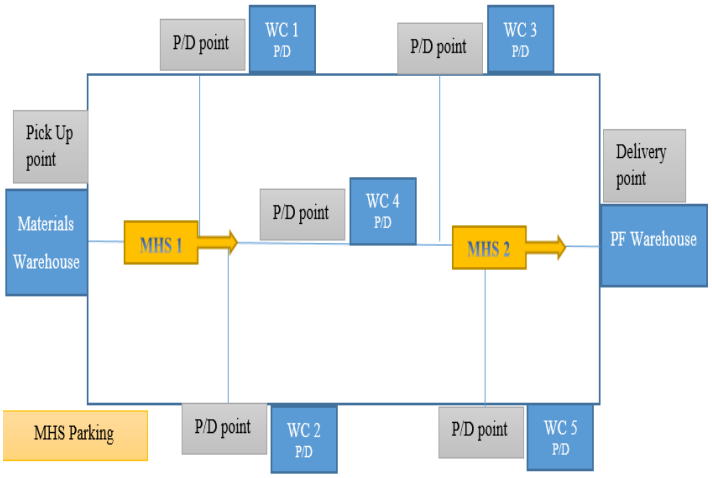


Figure 9. MS model including MHS

The MHS process, in this system, consist of four main steps:

- The unloaded MHS starts from the parking and moves to the pickup point to pick the load and goes to its destination (P/D point) related to the work center;
- The MHS stops to deliver the parts, picks up the new load if it exists;
- The loaded MHS moves to the next point whatever its storage area or work center, then the same operations will be repeated until the end of the procedure;
- MH returns to the parking.

The traveling time depends on: (1) the type of transporter (speed, loading time, unloading time, etc.); (2) the transfer network constraints like dispatching rule, orientation, distance between the departments etc.

5.3. Use of better Performance indicators

The performance indicators used by [21] are divided in two types:

- The main performance indicators are used to evaluate the global MS performance, in terms of tardiness, earliness and DD;
- Diagnostic performance indicators: they are used to resolve resource lack/surplus problems in each work center, which are represented as follows (Table 4):

Table 4.Indicators used by [21]

Indicators	Definition	Objective
Machine utilization rate (RUm)	Machine availability within a period of time	Maximize machine utilization
Labor utilization rate (RUop)	Labor availability within a period of time	Maximize labor utilization
Number of units in queue (WIP)	Total number of parts in queue and under process	Minimize WIP Parts in queue and parts in machine
Waiting time of units in queue	Total waiting time of parts in queue and under process	Minimize the waiting time of Parts in queue and parts in machines
Tardiness (Tr)	Time at which the task is completed after its date.	Minimize tardiness
Earliness (Er)	Time at which the task is completed before its date.	Minimize earliness

We can say that the performance indicators used by[21] are relevant but not sufficient to evaluate the performance of a MS including the MHS. The major shortcoming consists in neglecting the performance measure related to the MHS.

On the other hand, to improve SESA, it is necessary to measure the efficiency of the MHS, which can be characterized by several performance indicators.
In fact, previous research studies, in the field of fleet size, used different performance measures to analyze and evaluate the system performance. [18, 72-77].

Most of the indicators, related to MHS performance are:

- Vehicle utilization rate;
- Throughput rate;
- Vehicle waiting time;
- Vehicle blockage time;
- Empty vehicle traveling time;
- Total number of delivered loads;
- Number of units waiting to be transported;
- Waiting time of units to be transported;
- Availability/ unavailability of MHS.

5.4. Improvement proposal of the ES reasoning mechanism

The ES reasoning mechanism used by SESA is divided into two stages. The first stage aims at resolving machine selection problem without labor constraints (with a number of workers equal to the number of machines in each department). Then, the labor requirement problem resolution starts with a machine balanced MS.

According to our literature review, there has not been yet any expert system developed to select machines, labor and MHS simultaneously. In this work, we try to improve further ES reasoning mechanism in order to select not only the production resources (machines and labor), but also the material handling resources (trucks, pallets, etc.).To attain this objective, it is important to add another stage, in SESA, to resolve MHS fleet sizing problem (see Figure 10).

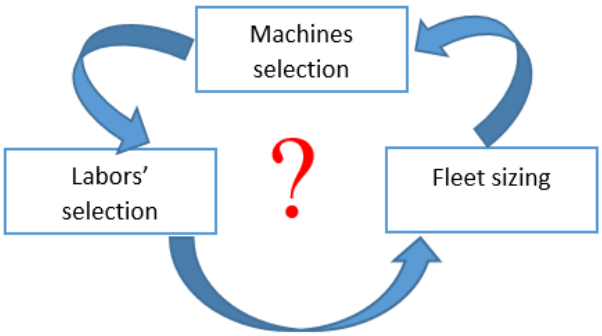


Figure 10. ES reasoning mechanism

The resolution of MHS fleet sizing problem requires new knowledge based rules related to the MHS in order to make the right decision of adding or deleting a MHS. The new stage must be in coherence with the other stages. We can take the example of the Lack and surplus of MHS resources. In fact, If the MHS utilization rate (ur) is less than the lower performance limits (supposed urmin=40%) there is a surplus of vehicles (the number of units waiting to be transported decreases but the MHS cost increases). In addition, If the MHS utilization rate (ur) is more than the higher performance limits (urmax=90%) there is lack of vehicles (the number of units waiting to be transported increases).

Example of the generic diagnosis production rule related to the lack of resources (ur>urmax):

MSS: Manufacturing system state.

If [GlobalPerformance: MSS= = OK] AND [D: ur>= D: urmax] Then D: problem=lack

The ES is a promising tool, as proved in previous section. It can be useful to various types of resource sizing problems. However, integrating an additional stage and finding an optimal order of three stages in the strategy structure is not an easy task. The feasible way to enhance SESA is to start with one problem, by giving a fixed number of resources for the others, then, using the solution of the first stage by solving the next problem, then, using the new information obtained from the second for solving the third problem. Each stage may have several iterations until satisfying results are obtained. In fact, it is necessary to identify the right ordering of selection problems because different ordering can provide different solutions. Unfortunately, the chained stages can address the complexity of the problem when dealing with a high-size problem. This approach becomes challenging to be applied.

Note that it is possible to combine other useful optimization approaches with simulation, such as the Artificial Neural Networks (ANN) Tabu algorithm, Genetic Algorithm, Data Envelopment Analysis (DEA), Regression Metamodeling etc. to identify the required number of resources.

6. Conclusion and Future Work

This paper discusses the originality and the evolution of SESA for sizing MS. The original version was created using static and deterministic tools to solve only the machine selection problems. Later research studies enhanced it by developing new stochastic and dynamic tools and enlarging its application domain by incorporating the labor selection problem. However, SESA is still limited, it neglects the MHS fleet sizing problem despite its influence on the machines and labor quantities. The provided literature survey indicates firstly that the MS sizing problem is continuously relevant due to its importance in ensuring the desired system performance. Secondly, the survey indicates that SESA is a promising approach for the MS sizing problem. The purpose of this study was to extend the application domain of SESA by integrating the MHS fleet sizing problem in the overall MS sizing task. The new SESA framework is thoroughly explained and its application and validation on actual case are proposed as future research.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgement

The authors gratefully appreciate the reviewers for their valuable remarks and suggestions, which have led to enhance this paper.

References

[1] I. Lajmi, W. Masmoudi, and M. Elleuch, "Importance of Handling Resources' Incorporation Into Simulation Expert System Approach," in 6th IEEE International Conference on Advanced Logistics and Transport (ICALT), pp. 139–143, 2017.

[2] Chtourou Hedi and Guillot Michel, "Une approche prescriptive simulée utilisant un système expert pour l'optimisation des ressources de systèmes

manufacturiers," Actes du 4éme congrés international de génie industriel, pp. 217–26, 1993.
[3] R. Arifin and P. J. Egbelu, "Determination of vehicle requirements in automated guided vehicle systems: A statistical approach," Production Planning & Control, vol. 11, no. 3, pp. 258–270, Jan. 2000.
[4] Chryssolouris, G, Lee, M, Pierce, J, and Domroese, M, "Use of neural networks for the design of manufacturing systems," Manufacturing Review, pp. 187–194, 1990.
[5] R. G. Askin and K. S. Chiu, "A graph partitioning procedure for machine assignment and cell formation in group technology" International Journal of Production Research, vol. 28, no. 8, pp. 1555–1572, Aug. 1990.
[6] F. O., H. Pierreval *, and D. Deflandre, "A simulation-based optimization approach to size manufacturing systems," International Journal of Production Research, vol. 43, no. 2, pp. 247–266, Jan. 2005.
[7] M. B. Yildirim, T. Cakar, U. Doguc, and J. C. Meza, "Machine number, priority rule, and due date determination in flexible manufacturing systems using artificial neural networks," Computers & Industrial Engineering, vol. 50, no. 1–2, pp. 185–194, May 2006.
[8] Z. Hajej, N. Rezg, and T. Askri, "Joint optimization of capacity, production and maintenance planning of leased machines," Journal of Intelligent Manufacturing, Oct. 2018.
[9] A. Rahman, S. Sarker, and M. T. Islam, "Simulating Cutting Line of a Furniture Industry," in 2018 International Conference on Production and Operations Management Society (POMS), pp. 1–7, 2018.
[10] A. Azadeh, S. Nazari-Shirkouhi, L. Hatami-Shirkouhi, and A. Ansarinejad, "A unique fuzzy multi-criteria decision making: computer simulation approach for productive operators' assignment in cellular manufacturing systems with uncertainty and vagueness," The International Journal of Advanced Manufacturing Technology, vol. 56, no. 1–4, pp. 329–343, Sep. 2011.
[11] A. Azadeh, S. M. Asadzadeh, and S. Tadayoun, "Optimization of operator allocation in a large multi product assembly shop through unique integration of simulation and genetic algorithm," The International Journal of Advanced Manufacturing Technology, vol. 76, no. 1–4, pp. 471–486, Jan. 2015.
[12] A. Ammar, H. Pierreval, and S. Elkosantini, "A multiobjective simulation optimization approach to define teams of workers in stochastic production systems," in 2015 International Conference on Industrial Engineering and Systems Management (IESM), Seville, pp. 977–986, Spain, 2015.
[13] S. Samant, V. K. Mittal, and R. Prakash, "Resource optimisation for an automobile chassis manufacturer through value stream mapping enhanced with simulation technique and constraint programming," International Journal of Industrial and Systems Engineering, vol. 28, no. 3, pp. 379–401, Jan. 2018.
[14] J. T. Lin, F.-K. Wang, and P.-Y. Yen, "Simulation analysis of dispatching rules for an automated interbay material handling system in wafer fab," International Journal of Production Research, vol. 39, no. 6, pp. 1221–1238, Jan. 2001.
[15] P.-H. Koo, J. Jang, and J. Suh, "Estimation of Part Waiting Time and Fleet Sizing in AGV Systems," International Journal of Flexible Manufacturing Systems, vol. 16, no. 3, pp. 211–228, Jul. 2004.
[16] D. Raman, S. V. Nagalingam, B. W. Gurd, and G. C. I. Lin, "Quantity of material handling equipment—A queuing theory based approach," Robotics and Computer-Integrated Manufacturing, vol. 25, no. 2, pp. 348–357, Apr. 2009.
[17] H. R. Sayarshad, "Using bees algorithm for material handling equipment planning in manufacturing systems," Int J Adv Manuf Technol, vol. 48, no. 9, pp. 1009–1018, Jun. 2010.
[18] F. F. Choobineh, A. Asef-Vaziri, and X. Huang, "Fleet sizing of automated guided vehicles: a linear programming approach based on closed queuing networks," International Journal of Production Research, vol. 50, no. 12, pp. 3222–3235, Jun. 2012.
[19] V. K. Chawla, A. K. Chanda, and S. Angra, "Automatic guided vehicles fleet size optimization for flexible manufacturing system by grey wolf optimization algorithm," Management Science Letters, pp. 79–90, 2018.
[20] V. Patel, J. Ashby, and J. Ma, "Discrete event simulation in automotive final process system," in Proceedings of the Winter Simulation Conference, vol. 1, pp. 1030–1034 vol.1, 2002.
[21] W. Masmoudi, H. Chtourou, and A. Y. Maalej, "Labor and machine sizing through a Simulation-Expert-System-based Approach," Simulation Modelling Practice and Theory, vol. 15, no. 1, pp. 98–110, Jan. 2007.
[22] H. Chtourou, W. Masmoudi, and A. Maalej, "An expert system for manufacturing systems machine selection," Expert Systems with Applications, vol. 28, no. 3, pp. 461–467, Apr. 2005.
[23] Wassim Masmoudi, H. Chtourou, and Aref Y. Maalej, "A simulation-expert-system-based approach for machine sizing of production systems," Jnl of Manu Tech Mnagmnt, vol. 17, no. 2, pp. 187–198, Feb. 2006.
[24] AshkanNegahban and Jeffrey S.Smith, "Simulation for manufacturing system design and operation: Literature review and analysis - ScienceDirect," Journal of Manufacturing Systems, vol. Volume 33, no. 2, pp. 241-261, Apr. 2014.
[25] D. R. Sule, Manufacturing Facilities : Location, Planning, and Design, Third Edition. CRC Press, 2008.
[26] S. M. Boorla, K. Bjarklev, T. Eifler, T. J. Howard, and C. A. McMahon, "Industry 4.0 – A challenge for variation simulation tools for mechanical assemblies," Advances in Computational Design, vol. 4, no 1, pp. 43-52, Apr. 2019.
[27] C. M. DeRienzo, R. J. Shaw, P. Meanor, E. Lada, J. Ferranti, and D. Tanaka, "A discrete event simulation tool to support and predict hospital and clinic staffing," Health Informatics J, vol. 23, no. 2, pp. 124–133, Jun. 2017.
[28] S. Almagooshi, "Simulation Modelling in Healthcare: Challenges and Trends," Procedia Manufacturing, vol. 3, pp. 301–307, Jan. 2015.
[29] L. Alonso et al., "CityScope: A Data-Driven Interactive Simulation Tool for Urban Design. Use Case Volpe," in Unifying Themes in Complex Systems IX, pp. 253–261, 2018.
[30] X. Song, Y. Wu, Y. Ma, Y. Cui, and G. Gong, "Military Simulation Big Data: Background, State of the Art, and Challenges," Mathematical Problems in Engineering, 2015.
[31] D. Mourtzis, M. Doukas, and D. Bernidaki, "Simulation in Manufacturing: Review and Challenges," Procedia CIRP, vol. 25, pp. 213–229, Jan. 2014.
[32] G. Popovics, A. Pfeiffer, and L. Monostori, "Generic data structure and validation methodology for simulation of manufacturing systems," International Journal of Computer Integrated Manufacturing, vol. 29, no. 12, pp. 1272–1286, Dec. 2016.
[33] A. Polenghi, L. Fumagalli, and I. Roda, "Role of simulation in industrial engineering: focus on manufacturing systems," IFAC-PapersOnLine, vol. 51, no. 11, pp. 496–501, Jan. 2018.
[34] A. Greasley, "Using simulation for facility design: A case study," Simulation Modelling Practice and Theory, vol. 16, no. 6, pp. 670–677, Jul. 2008.
[35] R. Zhang, S. K. Ong, and A. Y. C. Nee, "A simulation-based genetic algorithm approach for remanufacturing process planning and scheduling," Applied Soft Computing, vol. 37, pp. 521–532, Dec. 2015.
[36] D. Krenczyk, W. M. Kempa, K. Kalinowski, C. Grabowik, and I. Paprocka, "Production planning and scheduling with material handling using modelling and simulation," MATEC Web Conf., vol. 112, p. 09015, 2017.
[37] S. Yang, T. Arndt, and G. Lanza, "A Flexible Simulation Support for Production Planning and Control in Small and Medium Enterprises," Procedia CIRP, vol. 56, pp. 389–394, Jan. 2016.
[38] D. Ivanov, "Disruption tails and revival policies: A simulation analysis of supply chain design and production-ordering systems in the recovery and post-disruption periods," Computers & Industrial Engineering, vol. 127, pp. 558–570, Jan. 2019.
[39] W. Ye and F. You, "A computationally efficient simulation-based optimization method with region-wise surrogate modeling for stochastic inventory management of supply chains with general network structures," Computers & Chemical Engineering, vol. 87, pp. 164–179, Apr. 2016.
[40] Sławomir Kłos and Justyna Patalas-Maliszewska, "The Use of the Simulation Method in Analysing the Performance of a Predictive Maintenance System," in Distributed Computing and Artificial Intelligence, Special Sessions, 15th International Conference, pp. 42–49, 2019.
[41] S. Gallego García and M. García García, "Design and Simulation of Production and Maintenance Management Applying the Viable System Model: The Case of an OEM Plant," Materials, vol. 11, no. 8, p. 1346, Aug. 2018.
[42] J. Stoldt, A. Schlegel, and M. Putz, "Enhanced integration of energy-related considerations in discrete event simulation for manufacturing applications," Journal of Simulation, vol. 10, no. 2, pp. 113–122, May 2016.
[43] E. Romano and D. Iuliano, "A simulation/optimization approach to support resource allocation in a service firms," International Journal of Procurement Management, vol. 11, no. 1, p. 1, 2018.
[44] L. M. S. Dias, A. A. C. Vieira, G. A. B. Pereira, and J. A. Oliveira, "Discrete Simulation Software Ranking: A Top List of the Worldwide Most Popular and Used Tools," in Proceedings of the 2016 Winter Simulation Conference, Piscataway, NJ, USA, pp. 1060–1071, 2016.
[45] F. Hayes-Roth, D. Waterman, and D. Lenat, Building expert systems. 1984.
[46] M. H. Fazel Zarandi, S. Soltanzadeh, A. Mohammadi, and O. Castillo, "Designing a general type-2 fuzzy expert system for diagnosis of depression," Applied Soft Computing, Apr. 2019.
[47] O. T. Nkamgang, D. Tchiotsop, H. B. Fotsin, P. K. Talla, V. Louis Dorr, and D. Wolf, "Automating the clinical stools exam using image processing integrated in an expert system," Informatics in Medicine Unlocked, vol. 15, p. 100165, Jan. 2019.
[48] E. K. Tekez, "Development of an expert system for demand management process," International Journal of Computer Integrated Manufacturing, vol. 31, no. 10, pp. 970–977, Oct. 2018.
[49] Ali Ahmed Mohammed, K. Ambak, A. M. Mosa, and D. Syamsunur, "EXPERT SYSTEM IN ENGINEERING TRANSPORTATION: A REVIEW," vol. 14, p. 25, 2019.
[50] G. Baryannis, S. Dani, S. Validi, and G. Antoniou, "Decision Support Systems and Artificial Intelligence in Supply Chain Risk Management," in

Revisiting Supply Chain Risk, G. A. Zsidisin and M. Henke, Eds. Cham: Springer International Publishing, pp. 53–71, 2019.
[51] Jaskiranjit Kaur and Anjali Walia, "Dimensions of Expert System in Manufacturing Industry," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. Volume 3, p. 7, 2018.
[52] S. Singh, E. U. Olugu, and S. N. Musa, "Development of Sustainable Manufacturing Performance Evaluation Expert System for Small and Medium Enterprises," Procedia CIRP, vol. 40, pp. 608–613, Jan. 2016.
[53] K. Prasad, S. R. Maity, and S. Chakraborty, "Quality Function Deployment-based Expert System for Materials Selection," 2016.
[54] W. Hellmann, D. Marino, M. Megahed, M. Suggs, J. Borowski, and A. Negahban, "Human, AGV or AIV? An integrated framework for material handling system selection with real-world application in an injection molding facility," Int J Adv Manuf Technol, vol. 101, no. 1, pp. 815–824, Mar. 2019.
[55] R. Carreño, V. Aguilar, D. Pacheco, M. A. Acevedo, W. Yu, and M. E. Acevedo, "An IoT Expert System Shell in Block-Chain Technology with ELM as Inference Engine," International Journal of Information Technology & Decision Making, Jan. 2019.
[56] M. S. Amalnik, "Expert System Approach for Optimization of Design and Manufacturing Process for Rotary Ultrasonic Machining," vol. 11, no. 1, p. 13, 2018.
[57] H. Esen, T. Hatipoğlu, A. Cihan, and N. Fiğlali, "Expert system application for prioritizing preventive actions for shift work: shift expert," International Journal of Occupational Safety and Ergonomics, vol. 25, no. 1, pp. 123–137, Jan. 2019.
[58] C. F. Tan, S. N. Khalil, J. Karjanto, L. S. Wahidin, W. Chen, and G. W. M. Rauterberg, "An Expert Machine Tools Selection System for Turning Operation," IOP Conf. Ser.: Mater. Sci. Eng., vol. 88, p. 012044, Sep. 2015.
[59] C. Urrea, G. Henríquez, and M. Jamett, "Development of an expert system to select materials for the main structure of a transfer crane designed for disabled people," Expert Systems with Applications, vol. 42, no. 1, pp. 691–697, Jan. 2015.
[60] M. Ristić, M. Manić, D. Mišić, M. Kosanović, and M. Mitković, "IMPLANT MATERIAL SELECTION USING EXPERT SYSTEM," Facta Universitatis, Series: Mechanical Engineering, vol. 15, no. 1, pp. 133–144, Apr. 2017.
[61] M. Soleymani and M. O. Nejad, "Supply Chain Risk Management using Expert Systems," International Journal of Current Engineering and Technology, vol. 8, no. 04, 2018.
[62] P. S. Kang and R. S. Bhatti, "Continuous process improvement implementation framework using multi-objective genetic algorithms and discrete event simulation," Business Process Mgmt Journal, Sep. 2018.
[63] A. Burduk, K. Grzybowska, and G. Kovács, "Hybrid model of an expert system for assessing the stability of a production system," LogForum, vol. Vol. 14, no. 4, 2018.
[64] A. Singh, D. Philip, J. Ramkumar, and M. Das, "A simulation based approach to realize green factory from unit green manufacturing processes," Journal of Cleaner Production, vol. 182, pp. 67–81, May 2018.
[65] J.-P. Arnaout, G. Arnaout, and J. E. Khoury, "Simulation and Optimization Of Ant Colony Optimization Algorithm For The Stochiastic Uncapacitated Location-Allocation Problem," Journal of Industrial and Management Optimization, pp. 1215–1225, Oct. 2016.
[66] K. Nagadi, L. Rabelo, M. Basingab, A. T. Sarmiento, A. Jones, and A. Rahal, "A hybrid simulation-based assessment framework of smart manufacturing systems," International Journal of Computer Integrated Manufacturing, vol. 31, no. 2, pp. 115–128, Feb. 2018.
[67] D. Stefanovic, N. Stefanovic, and B. Radenkovic, "Supply network modelling and simulation methodology," Simulation Modelling Practice and Theory, vol. 17, no. 4, pp. 743–766, Apr. 2009.
[68] B. Hernández-Vera, A. A. A. Lasserre, M. G. Cedillo-Campos, L. E. Herrera-Franco, and J. Ochoa-Robles, "Expert System Based on Fuzzy Logic to Define the Production Process in the Coffee Industry," Journal of Food Process Engineering, vol. 40, no. 2, p. e12389, 2017.
[69] A. Azadeh, H. Tohidi, M. Zarrin, S. Pashapour, and M. Moghaddam, "An integrated algorithm for performance optimization of neurosurgical ICUs," Expert Systems with Applications, vol. 43, pp. 142–153, Jan. 2016.
[70] H. Behret, B. Öztayşi, and C. Kahraman, "A Fuzzy Inference System for Supply Chain Risk Management," in Practical Applications of Intelligent Systems, pp. 429–438, 2012.
[71] A. Hadi-Vencheh and A. Mohamadghasemi, "A new hybrid fuzzy multi-criteria decision making model for solving the material handling equipment selection problem," International Journal of Computer Integrated Manufacturing, vol. 28, no. 5, pp. 534–550, May 2015.
[72] B. Mahadevan and T. T. Narendran, "A hybrid modelling approach to the design of an AGV-based material handling system for an FMS," International Journal of Production Research, vol. 32, no. 9, pp. 2015–2030, Sep. 1994.
[73] B. M. Beamon and V. C. P. Chen, "Performability-based fleet sizing in a material handling system," The International Journal of Advanced Manufacturing Technology, vol. 14, no. 6, pp. 441–449, Jun. 1998.
[74] Hung, P. C. and Liu, F. H, "Estimation of the fleet size for a multi-load automated guided vehicle system," Proceedings of the National Science Council, Part A: Physical Science and Engineering, pp. 244–253, 2001.
[75] S. Rajotia, K. Shanker, and J. L. Batra, "Determination of optimal AGV fleet size for an FMS," International Journal of Production Research, vol. 36, no. 5, pp. 1177–1198, May 1998.
[76] K.-H. Chang, A.-L. Chang, and C.-Y. Kuo, "A simulation-based framework for multi-objective vehicle fleet sizing of automated material handling systems: an empirical study," Journal of Simulation, vol. 8, no. 4, pp. 271–280, Nov. 2014.
[77] R. Wibisono, T. J. Ai, and D. R. Yuniartha, "Fleet Sizing of Automated Material Handling Using Simulation Approach," IOP Conf. Ser.: Mater. Sci. Eng., vol. 319, p. 012030, Mar. 2018.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 224-233 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Fuzzy Simulation of Historical Associative Thesaurus

Olena Basalkevych[*,1], Olexandr Basalkevych[2]

[1]*Danylo Halytsky Lviv National Medical University, Department of Medical Informatics, 79005, Ukraine*

[2]*Product Engineering and Development Globallogic Inc., 79016, Ukraine*

ARTICLE INFO

Article history:
Received: 03 June, 2019
Accepted: 13 September, 2019
Online: 08 October, 2019

Keywords:
fuzzy associative field
stimulated fuzzy associative field
fuzzy associative word structure
fuzzy associative map
fuzzy historical associative thesaurus
associative antipodes
associative experiment
Osgood's semantic differential
fuzzy associative differential
fuzzy associative stratification
Older Scots.

ABSTRACT

Research is dedicated to fuzzy reconstructions in the diachronic associative linguistics, and encompasses several simulation practices. The principles of an associative memory model and fuzzy computing with words are adopted in the working methodology. The evolved mathematical model reanimates the older patterns of Scottish associative map by means of fuzzy logic. The compiled fuzzy associative fields are connected by epidigmatic relations represented by fuzzy associative word structures in a fuzzy associative map constituting the grounds of the proposed thesaurus. The simulation of the historical associative thesaurus enables extended associative scrutiny: tracing associative antipodes, reconstructing an associative experiment for any map associate, finding words' similarities and distances, as well as associative differentiating and stratifying the Older Scottish lexis. The special interest represents the fuzzy associative differential, an electronic tool for estimating the lexis through its stable syntagmatic associates. All the simulation processes are fuzzy initiated and quantified. The proposed practice in its final configuration has no precedents, representing the novelty of theoretical interpretation and empirical application. The proposed terminology is mainly introduced by the researchers.

1. Introduction

Ethnoconsciousness of older speakers is the matter of research in the represented work. Associative maps of the modern ethnoses are mainly grounded on an associative experiment, which allows ascertaining the lingual associations through collecting the reactions of respondents to stimuli, the practice that is absolutely impossible for literal realization at the older language-states. Logically, to reconstruct the map and subsequently to simulate an associative experiment in the diachronic psycholinguistics seems real under the study of lexical relations present in the synchronic associative map. Such relations could be artificially reconstructed, quantified and canonically combined to reanimate the mental lexicon and all the related issues for the periods more or less remote in time.

In our research, the capability of fragmental reconstructing and activating the mental lexicon of Older Scots is realized with fuzzy modelling. The first turning point of the research is compilation of the diachronic synonymic sets of Older Scottish adjectives. The source of compilation is the Dictionary of Older Scottish Tongue (DOST) [1], an explanatory dictionary, providing the English description to Older Scottish lexis. The incentive is a synonymic set of English adjectives, any of which becomes an explaining semantic feature for the extricated Older Scottish synonyms. All those synonyms are gathered into a set, and ordered chronologically. Such sets are characterized by the connotative, stylistic and ethnic variability. The second decisive moment is the concluding of associativity nature of the compiled sets. The fuzzification of the sets is enabled by Zipf's regularity allowing to treat the diachronic measure, calculated for any set synonym, as that of associativity. In our research the compiled fuzzy sets are called the fuzzy associative fields (FAFs). The revealed fuzzy associative word structures (FAWSs) connect FAFs into the network, called an associative map or associative lexicon of Older Scots. Another associative relationship – antonymy – is brought into the structure, joining any FAF with its antonymic one. Thus a two-wing associative map is evolved at the stage. The fuzzy marking of the associative map allows us to

[*]Olena Basalkevych, 12 Zelena str., Lviv, 79005, +380955365348, olena.basalkevych@gmail.com

run the evolved network, realizing multitask simulation of the mental Older Scottish lexicon. Thus the first type of the run enables finding the associative chains through positive and negative wings of the map, called associative antipodes.

The following iconic associative relationship – syntagmatic one – attaches the available in the dictionary collocates to any adjective-knot of the network. Then the associative experiment in its nominal format is ready to be reproduced, bringing to any adjective of the map the words from its fuzzy associative structure; all existing till the moment associates within the major field; time antonym with its dominant and, finally, an available collocate. All the entries are collected in one fuzzy set, called a stimulated fuzzy associative field (SFAF). The simplest fuzzy metrics of similarity and distance could be found only within the same FAF, since they are not context- and turn-sensitive. The junction of the two antagonist-fields is regarded as a fuzzy differential, which enables the differentiation of the available Older Scottish lexis throughout the network. It gathers for any word the fuzzy subset of all its syntagmatic collocates available within the associative fragment. The new recommended method introduces into a traditional Osgood's semantic differential the expected denotative meanings. The electronic tool, created for the purpose, we call a fuzzy associative differential. The fuzzy stratification of Older Scottish lexis is another issue of the same approach. The electronic resource realizing all the facilities for the activated part of the associative map is called a fuzzy historical associative thesaurus (FHATh). All the techniques are grounded on the classical principles, notified in the following chapter, but represent an independent system, compiled in the series of previous author's works [2-4], mentioned throughout the paper. The absolute novelty of the practice lies in its possibility to imitate the older mental lexicon patterns on the basis of the proposed row of reconstructions, which was not applied in the diachronic linguistics before. The created electronic resource completely corresponds to the principle of computing with words: the user is pressing a word in the map and software returns the estimating phrase (supported with illustrating numbers) in response [5].

The classical sources whose concepts make the background of the developed mathematical model are analyzed in the following Chapter 2; and the comparison with the resonant works is realized afterwards (Chapter 3). The stages of development of the mathematical model are described in Chapter 4. The realization of the model resulted in different research techniques is explained in Chapter 5. The software packet and its principles of computing with words are described at the very end of the chapter. Finally, the received results are discussed (Chapter 6), and general conclusions are drawn (Chapter 7).

2. Report on Scientific Sources

2.1. Associative thesaurus of J.Kiss

One of the cotemporary and highly professional patterns of an associative map, reflecting the associative structure of words, was developed in 1972 under the aegis of British Medical Research Council by J.R. Kiss in his "An associative thesaurus of English", which is defined as a word association network represented by an oriented linear graph with labelled arcs, where the nodes are words and labels on the arcs give estimates of stimulus-response probabilities. As to its organization: "Two different approaches to the semantic structuring of the thesaurus data can now be distinguished. One is aimed at the discovery of semantic components for a word (covering all its word senses). The other is aimed at the discovery of groups of words which are relevant to each other because the concepts they label go together (although may not be similar to each other)" [6]. The concept becomes the first foundation of the proposed work. Our oriented graph of a network configuration called a fuzzy associative network (FAN) is organized following both processing approaches, the latter is well represented by FAFs gathering Scottish words whose main semantic components "go together" in synonymic sets of English adjectives, and the former completely corresponds to FAWSs.

2.2. Age-Frequency Regularity of G. Zipf

The second foundation is the regularity of G.K. Zipf stating the relationship between frequency and age of a word: "Hence, in sum, we have found not only a direct relationship between relative age and relative frequency, but also an inverse relationship between relative age and relative size in words of the same frequency" [7]. His conclusion on the English lexis time-age correlation states that the rank of the most frequent modern English words contains mostly the words originated from Old English [7]. In our research, Zipf's statement allows us to treat a diachronic measure as a frequency measure and therefore as an associativity one.

2.3. Fuzzy Logic

The state of affairs actualizes the third foundation of the work that is fuzzy logic, which shares the mathematical apparatus ready to fix the partial state of truth of a statement in our mind or speech. The honoured founder of the theory Lotfi Zadeh was working with the human language when fuzzy logic emerged, the logic implying the numerousness of partial truths on the way from nonsense to absolute. Lotfi Zadeh argued that probability lacks sufficient "expressiveness" to deal with uncertainty in the natural language. According to [8], in his "Computing with Words" Zadeh writes: "Humans have many remarkable capabilities. Among them there are two that stand out in importance. First, the capability to converse, communicate, reason and make rational decisions in an environment of imprecision, uncertainty, incompleteness of information and partiality of truth. And second, the capability to perform a wide variety of physical and mental tasks without any measurements and any computations. In large measure, Computing with Words is inspired by these remarkable capabilities."

A linguistic variable refers to the set of values more or less representing the emerged truth. The degrees of such a representation, or the measures of truth partiality, are caught by *membership functions*, normally, real numbers lying in the interval from 0 to 1 inclusively. The set is called a *fuzzy set*.

The formal definition: Let E be a countable or uncountable set and x be an element from E. Then *a fuzzy subset A of the set E* is characterized as a set of the ordered pairs $\{(x, \mu_A(x))\}, \forall x \in E$; where $\mu_A(x)$ is a characteristic fuzzy function (or a membership function) that takes its values in totally ordered set M or within [0;1], and indicates the level, or degree, of the element x belonging to the subset A. A linguistic variable alters in names, taking them from the set of verbal (word) characteristics,

representing all the sorts of it. Normally, such a set is called *a term set* [9, 10].

The fuzzification of our structure starts from the associativity measure that corresponds to a membership function, reflecting the degree of associativity of field members with a dominant. In short terms, Zipf's regularity enables fuzzy simulation of the diachronic associative thesaurus.

The fuzzy metrics of distance and similarity borrowed from the same theory [9] are available for all the elements of the represented associative map fragment.

Granularity in fuzzy logic implies dealing with data points within a granule as a whole rather than individually [11]. In our work it is used in consolidation phase while differentiating Older Scottish words. Available word-characteristics with their fuzzy estimates are grouped into six concept sets or granules with deriving the general probability and fuzzy measures for any granule. Fuzzy stratification is another way to address to the point data through the stratum which contains them [12]. The approach seems to be very profitable in the pay of the historical lexicography, naturally laying its verbal wealth layer over layer.

Thus the mathematical model of the represented work is extendedly grounded on fuzzy logic.

2.4. Osgood's Semantic Differential

The constructive element of FHATh is a junction of two antonymic FAFs which is constructed and recommended by our research to be used as a fuzzy associative differential (FAD). A FAD could be treated as a fuzzy diachronic modification of Ch. Osgood's semantic differential. FAD represents a lot of direct associates to different qualities, while the scales of the original Osgood's differential contain the intensity markers of these qualities. Preferably, literary meanings, since all of them are taken from the dictionary, are used in the scales of FAD, bearing classical denotative meanings with different connotative shades. Therefore, Osgood's conflict of 'connotative emotions' versus 'denotative precision' and the symbolic character of his differential scale-intervals [13] are moderated in our development. Meanwhile, the impartiality of Osgood's intensity markers whose applicability span could not be overestimated lacks in FAD. The most popular scales of Osgood's differential are Good-Bad, Strong-Weak and Active-Passive, representing so-called cross-cultural universals of Evaluation, Potency and Activity of the considered words. Our research adopts the universals for its measuring purposes.

2.5. Associative Network Memory Model

The principle of the network performance is comparable with "spreading activation" of an associative network memory model: when the concepts of interest are activated, "excitation" penetrates through the links to the associated concepts. More frequent in the past associations are stronger and easier to be excited. If the emitted excitation is potent enough to activate some of those associations and surpass the threshold, "activation is spread" and new concepts are brought to mind. The associative memory structure is highly metaphorical but still the only manner of simulating the physiology of the brain [14]. Normally, the network is evoked episodically in response to the stimulus, visualizing the associative map fragments which combine logogens as images of words, containing the information on the word valency, including stable associations. In the present work, the map is also activated episodically. Information on an associativity level of the map words as well as the smart orientation ability is hidden in the fuzziness of the network patterns.

3. Comparison Analysis

3.1. Fuzzy Hindi WordNet

In line with the proposed research, the fuzzy interpretation of Hindi WordNet [15] is realized with a fuzzy graph, and represents an ideal comparison pattern with similar underlain principles and inner graph morphology but a different fuzzification method. Our original work [2], which appeared earlier than the regarded comparison source, concentrated more on the paradigmatic and epidigmatic associativity of the proposed map. The most salient and fundamental difference of Fuzzy Hindi WordNet lies in treating the fuzzy graph vertices, synonymic sets, so-called synsets of nouns, verbs, adjectives, and adverbs as the sets of open-class words where the strength of every node is always 1 [15]. In our study the associative fields of adjectives are represented by fuzzy sets, employing the inherited vagueness of linguistic variable values that are normally represented by adjectives as degree-measurers [16]. Evidently, two works differ much by the calculus of membership functions: the classical expert opinions used in Fuzzy Hindi WordNet [15] seem to be hardly reachable for older languages, therefore we reasonably adapt a diachronic measure for the purpose. The weighed in the mentioned manner semantic and lexical connections in [15] are organized among synonymic sets and words, respectively, through association, hypernymy, hyponymy, meronymy, holonymy, antonymy, entailment, troponymy, gradation, and causative ties, whilst in [2] the epidigmatic relations are combining the different associative fields through the entries of the same words. As to the antonymic relation, in our case it immediately connects diachronic antonyms and the whole antonymic fields, containing them. The junction is realised through the antonymy of their dominants, following the proved associativity of antonyms [17]. The goals of fuzzy upgrade also diverge, since we represent a fuzzy reconstruction device whilst [15] is more about fuzzy metrics allowing to extricate the context word values in the modern Hindi discourse and text. As to the research sway, our work is rather an illustrated recommendation to a new type of historical thesaurus while Fuzzy Hindi WordNet represents an accomplished modern lexicography development.

3.2. Fuzzy Neural Networks

In the work, the morphology of the mentioned network is sophisticated in stipulation but simple in performance, and probably with much higher structural accuracy could be realized by means of a fuzzy neural network. It is equipped with fuzzy-valued connection weights and morphological neurons good at logical operations of conjunction and disjunction or dilation and erosion in terms of [18] among the others. All the more argument is that a neural network seems to be the most appropriate for simulating the associative reasoning [19]. Nevertheless, at the current stage of the research the proposed in the article fuzzy

model suffices completely the purposes of the set lexicological research.

4. Mathematical Model

The exposed interdisciplinary correlation of facts and dogmas flows out into the model serving the diachronic associative linguistics with fuzzy logic. The principles were originally conceived in the work [2], and developed in the following works, including [3, 4]. There could be discerned several stages in the evolution of the model.

4.1. Compilation of the Diachronic Synonymic Sets of Older Scottish Adjectives

The technique starts from the English synonymic adjectival sets taken from a dictionary of synonyms of Oxford Learner's Thesaurus (Oxford Thesaurus 2008) [20] or WordNet (WordNet 2006) [21]. They become the search stimuli in the DOST: the resulting Scottish adjectives contain them as defining semantic features. The reactions to all the English set adjectives are compiled into the set of Older Scottish synonyms. Their frame registers the first citations they were mentioned in, or so-called diachronic text prototypes [22], the year of the citations, or so-called advent time, their source, encyclopedic meanings and etymological notes. Synonyms are ordered in a chronological order, and any of them is attributed with a diachronic measure, calculated due to the formula:

$$W_i^{(y)} = \frac{y_i - y_{\min}}{y_{\max} - y_{\min}}\left(\frac{1}{n+1} - 1\right) + 1, \qquad (1)$$

where ($y_{\max} - y_{\min}$) is a diachronic range of a set, ($y_i - y_{\min}$) is a distance to the appearance of the concrete word in a set, and n – the number of words in a set [23]. The organization and type of the sets imitate the Historical Thesaurus of the Oxford English Dictionary [24]. Primarily, the sets are called diachronic synonymic sets, and there are 100 of them in our database. They result in very heterogeneous collections of words affiliated by the same or comparable semantic features but reflecting the notable wideness of connotative, stylistic and ethnic coloration, which evidently leads them out of ordinary synonymy.

4.2. Associativity Nature of the Constructed Diachronic Synonymic Sets

The Dictionary of Older Scottish Tongue (DOST) is a resource most readily associated with Older Scots knowledge in people's mind, according to the dictionary definition given by Budanitsky [13]. The historically forged layers of lexis are stored there, keeping the primary syntagmatic associations reflected in numerous collocations and sentences, which contribute to the secondary paradigmatic associations [25]. Since all the words are dated, and represented with diachronic text prototypes, the historical dictionary like the DOST is expected to be the people's memory enfolded. The organization of memory is hypothetically comparable with an associative map as is mentioned in Chapter 2.5.

Therefore, the set goal of the research is to unfold fragmentally the associative map of Older Scots from the DOST. The paradigmatic collections within the received diachronic synonymic sets reflect the associative relationship, since *the process of their compilation is regarded as a discrete associative experiment with historical reaction* (the traditional associative experiment counting numerous reactions to a stimulus is called a discrete associative experiment with prolonged reaction [14]). The English synonyms, the DOST search items, play the role of stimuli. Thus at the stage, we rename the diachronic synonymic sets into synonymic associative fields.

4.3. Synchronization of the Associative Structures

The projection of the diachronic synonymic sets into the synchronic plane is realized by the mentioned in Chapter 2.2. Zipf's *age-frequency* correlating regularity, which states that the oldest words of the lexicon are the most frequent ones. On the other hand, the first words given during an associative experiment are the characteristics most frequent in mind, according to [14, 26]. Both facts directly influence the nature of the diachronic measure (1): it is converted into a frequency measure, and then in an associativity one. The hypothetical syllogism leads to the treatment of diachronic coefficients (1) as values of associativity of the field members with their dominant.

4.4. Fuzzification

Following the tenets of Chapter 2.3., E is a huge countable set of all the Scottish words (or adjectives only). From E, we could discern the fuzzy subsets A_j of words rendering specific semantic features or dominants, represented by English adjectives, which give the names to A_j. Membership functions or the degrees of reflecting these semantic features are completely expressed by associativity measures (1). Consequently, a Dominant is a linguistic variable and takes the values from its term set: *Dominant={Good, Bad, Strong, Weak, Active, Passive, Flexible, Inflexible, Brave, Cowardly, Intelligent, Stupid, Large, Small, Lucky, Unlucky,…}*. Altogether it contains 100 name-values (j=1:100). In our research, fuzzy subsets A_j are called *fuzzy associative fields* [2-4] or FAFs hence. For example, FAF *Active*= (douchty/1, fere/1, deliver/1, strenthy/1, licht/1, forthirwarde/0.99, unswere/0.87, agile/0.87, fery/0.87, active/0.78, wirkand/0.7, operative/0.66, lifly/0.63, tait/0.63, ȝape/0.63, restles/0.63, frak/0.63, trig/0.63, laborious/0.59, nocturnall/0.59, slepry/0.53, liver/0.52, throuche/0.49, gangand/0.46, actual/0.43, feat/0.42, ferdie/0.36, ȝauld/0.32, suddan/0.3, spritly/0.12, vigent/0.03).

From the totality of FAFs, we can extricate an associative word structure [2-4], epidigmatic relationship, stating the level of word associativity with different Dominants. We call it a *fuzzy associative word structure* [2-4] or FAWS hence. The example of FAWSs: Douchty=(Active/1, Vigorous/1), Agile=(Active/0.87, Vigorous/0.87), Unswere=(Active/0.87, Vigorous/0.86, Helpful/0.83).

Commenting the given instances, the variability of belonging for words Douchty, Agile and Unswere is not large, we observe the same Dominants with high associativity degrees in all the three cases, which states the intension nature of the mentioned semantic features.

4.5. Fuzzy Association Map

The unfolding of the associative map hidden in the historic dictionary starts from the connection of the evolved fuzzy

associative relations into the structure corresponding to the requirements of an associative thesaurus of English by J.Kiss [6]. The structuring approaches are well described in Chapter 2.1. and are completely sufficed by the combination of the developed FAFs and FAWSs into a map. One more strong associative relation proved by an associative experiment is an antonymic one [17, 27]. In our visualization, technically saying, we connect three positive FAFs Good, Strong, and Active with their antonymic ones Bad, Weak, and Passive, creating two wings of the *fuzzy associative map* or FAM [3, 4]. Any map wing contains a lot of between-field connections, combining the same words located in different fields. Some of the words and connections are distinctly represented in Figure1. The between-field connections themselves introduce into the map the epidigmatic relationship and are given by FAWS.

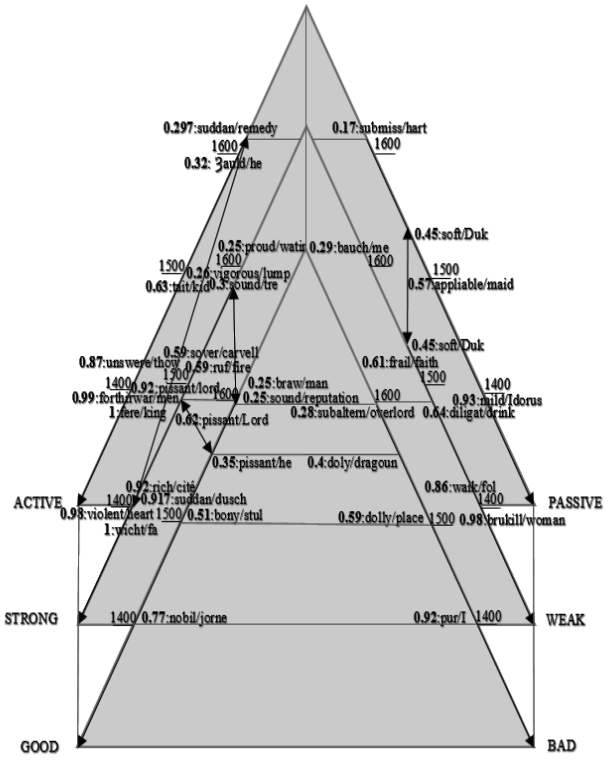


Figure 1: Fragment of FHATh.

The given fragment of a fuzzy associative map of Older Scottish adjectives covers the scales of the cross-cultural universal involved in the associative techniques of lexis Potency (Good-Bad), Activity (Active-Passive) and Evaluation (Good-Bad) ascertaining. Far not all the associates are visualized there: technically complete scales are represented by software forms.

5. Simulating Techniques

The compiled map constitutes the ground for FHATh. In a final condition it is lined up with syntagmatic associations: any adjective of the set is paired with noun or pronoun taken from the diachronic text prototypes, represented by the DOST. The diachronic text prototype is the oldest quotation registered in the literature, containing the considered word. Following Zipf's regularity, we proved that the oldest word is the most frequent one and therefore contains the greater associativity strength. The same concerns the gained by the word syntagmatic relations, documented in the diachronic text prototype, which are the most prototypical and thus the most associatively stable. (Find the collocations in Figure 1.) Their fuzzy strength due to propagation acquires the fuzziness of the adjective.

The associativity of synonymic, antonymic and syntagmatic relations is supported by Deese in his «The structure of associations in Language and Thought», described in [17]. So the complete associative map is justified and thus ready to run. The activation of the associative map corresponds to the "spreading activation", caused by the "fired" stimulus, and is well-directed, since the map is represented by an oriented graph of a network configuration. The fuzziness plays its selective function in the process of network navigation. An associative map activated in the manner is called by our research a *fuzzy associative network* or FAN, and the process of its target activation is hence named the simulation of a target.

5.1. Associative Antipodes

The first target is an associative environment. Simulation of the associative environment ends up finding two associative chains through positive and negative map wings, called associative antipodes. To find the associative environment to the chosen on the map word, algorithm traces this word FAWS, detecting the word-place with the maximum membership function, after falls to the dominant of the found word as one with the maximum associative measure or membership function, and repeats the same scenario for this dominant till no more connections could be revealed, finding the n-association chain to a word-stimulus at the 2n-th step. The fuzziness of a chain acquires the minimum membership function of the passed places, according to conjunction in fuzzy logic:

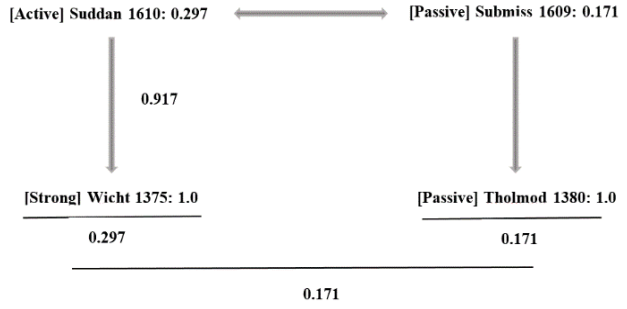
$$\mu(\mathrm{x}_1 \cap x_2) = \min[\mu(x_1), \mu(x_2)]. \quad (2)$$

Then the same figurant is connected with its time-antonym, and the run is repeated within the opposite map wing till at the 2m-th step it reaches the m-association chain with the minimum membership function. Both chains constitute the n/m associative environment for the chosen word, or associative antipodes. The activated in the manner map we call *2 run fuzzy associative network* (2RFAN) [4]. In Figure 1, two pairs of antipodes are traced according to the algorithm. They are graphically derived in Figure 2.

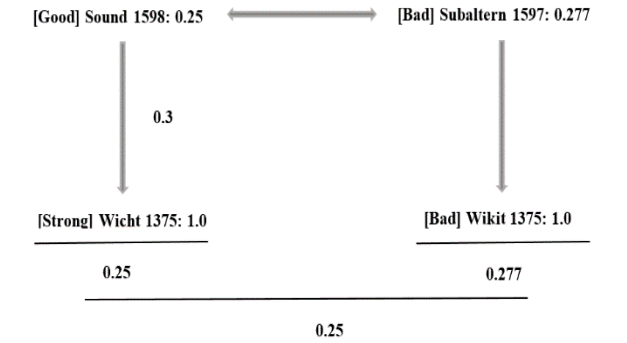
The program realization of any thesaurus function, including the tracing of an associative environment, meets the principle of computing with words: a user is pressing a word in the associative map (Figure 3) and receives the evaluating phrase about the word's relations. The phrase is constructed on the basis of fuzzy values involved in estimating the associativity level or degree of belief of the received results (Figure 4).

The console (Figure 4) states that the positive associative environment for the word Suddan from the field Active is supposed to be Wicht from the field Strong with Low degree of belief (0.3). Its negative associative environment is supposed to be Submiss from the field Passive and the dominant Tholmond

from the same field, this time with Very low degree of belief (0.17). In its turn, the positive environment for Sound from Good is also Wicht from Strong but with lower degree of belief (0.25), and the negative antipode here is Subaltern from Bad and Wikit from Bad with Low degree of belief (0.28). The data support the graphics in Figure 2.



(a)



(b)

Figure 2: Associative antipodes: (a) [Suddan/Wicht]-[Submiss/Tholmod]; (b) [Sound/Wicht]-[Subaltern/Wikit]

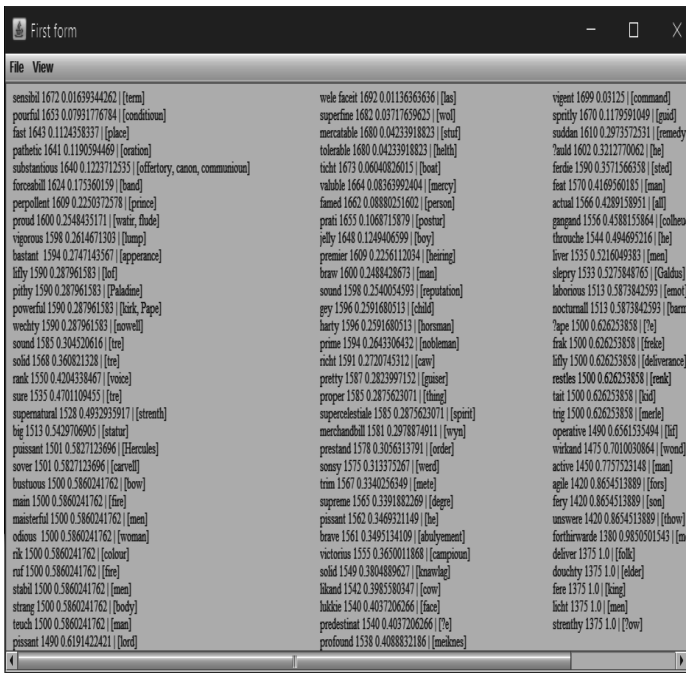


Figure 3: Map for selecting the words and realizing the simulation techniques.

In both cases we observe one semantic shift through positive map wing and no one but falling to the dominant in the negative wing. The same links could be singled out in Figure 1.

The found associative ramifications reveal the deepness of the associative map. On the limited fragment it shows quite a good coherence tendency.

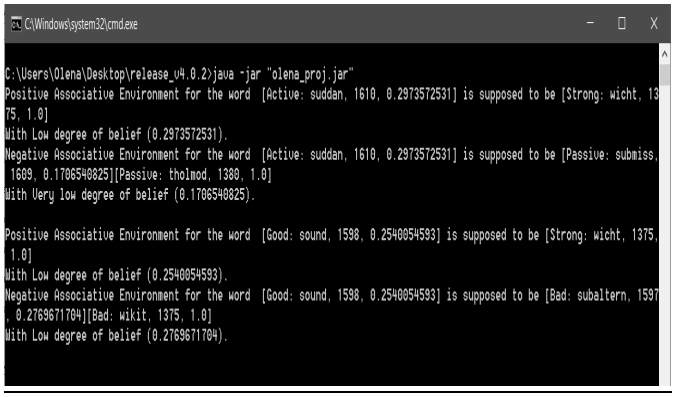


Figure 4: Console with results of tracing an associative environment for the word Suddan from the field Active, and Sound from the field Good.

5.2. Associative Experiment

According to the revealed regularities of the real associative experiment [14, 17, 25, 26], its schematic reproducing includes associative structure of the word (epidigmatic components), and from the field where its associativity is maximal: older synonyms of the word (paradigmatic components), collocates (syntagmatic components), time antonym and the dominant of a time antonym (paradigmatic components). All the entries are grouped into one fuzzy set or a *stimulated fuzzy associative field* or SFAF. Herewith, the dominant of the most associative field is entering the set with its proper fuzzy value 1 for subjective reasons of immersing in its mental plan. The membership function of a collocation is suggested by the adjective fuzziness. In such a case, an experiment could be treated as independent. The context-dependent experiment extracts the infield associations from the FAF of word localization. For instance, on pressing Pissant at the scale Good, within the framework of the independent associative experiment the algorithm traces all dominants Pissant belongs to [wicht/0.62, quhite/0.35]. The strongest association is observed in the FAF Strong, so an algorithm flexibly changes the field and extracts all the older synonyms, collocate Lord and time antonym Diligat with its dominant Tendir from FAF Weak, adds to the whole "community" the remaining dominant Quhite from the "calling" FAF Good, and orders the received fuzzy set. The program release of an associative experiment simulation for the stimulus Pissant is in Figure 5:

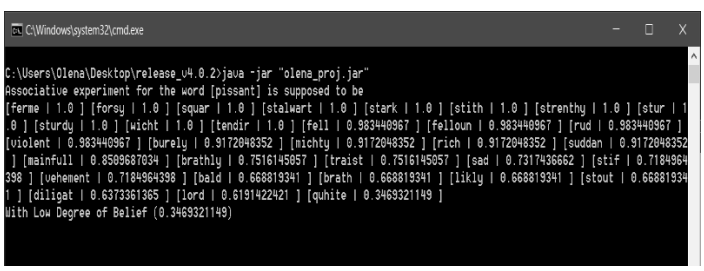


Figure 5: Console with the results of simulating an associative experiment for the stimulus Pissant.

The stability of the given SFAF or the associative reaction to the stimulus is read from the minimum membership function of its entries, normally that of the last element of the ordered SFAF. In our case, it is 0.35: with Low degree of belief the Older Scot would react to the stimulus Pissant in the proposed manner. If one needs more trustful result, the set should be cut at the required fuzziness level, expelling the newest and thus the rarest words.

5.3. Distance and Similarity

So far the infield trivial fuzzy metrics have been employed by FHATh to quantify the similarity and distance between associates. Thus the distance between associate x_i and associate x_j is

$$D = \left|\mu(x_i) - \mu(x_j)\right|, \quad (3)$$

where $\mu(x_i)$ and $\mu(x_j)$ are their membership functions or associativity measures, respectively, and their similarity is

$$S=1-D. \quad (4)$$

The classical fuzzy metrics' formulas are employed, and since they are not semantic-shift sensitive, we apply them only within a field. The between-field associativity metrics seem to be more complicated, and already regarded associative antipodes bring the maximum information on the direction and value of existing associativity among different fields. The program release for Distance and Similarity calculus for word-associates Valuble and Victorius from FAF Good is given in Figure 6:

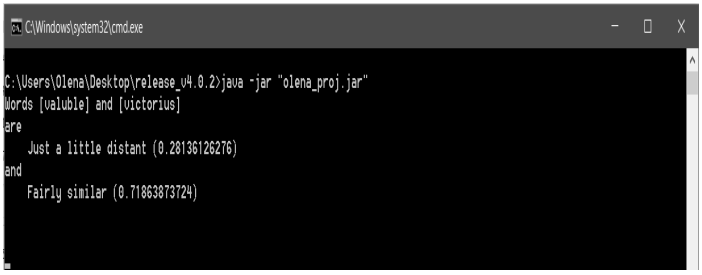


Figure 6: Release of fuzzy distance and similarity calculation for associates Valuble and Victorius from the FAF Good.

The Distance between Valuble and Victorius is rather small 0.28, that is why the software concludes that they are just a little distant. Therefore, the Similarity is quite high 0.72, with the conclusion that they are fairly similar.

5.4. Fuzzy Associative Differential

The function of the proposed by authors FAD is to provide as many associations to the considered quality as possible to describe it naturally and softly. The junction of antonymic FAFs serves the best for the purpose [3, 4]. All possible subtleties are gathered there and measured with associativity degrees. Three universals of Evaluation, Potency and Activity are depicted in Figure1: the space allows us to provide just sectional semantic values, the completeness of them is visible in the program forms (one of them is given in Figure 3). The scales are ready to serve any lexical request, and the tiniest shades of the quality could be communicated by them to the regarded word. In the case of the older language patterns, the service reflects the stable language collocations, provided by the diachronic text prototypes, and therefore treated as the most strengthened in time. So we gather all the possible characteristics for the collocate, regarded in the map, throughout all the six differential's scales and consolidate them into six fuzzy sets, or six granules in terms of fuzzy logic, mentioning the proper membership function $\mu_{Granule}$ and probability $P_{Granule}$ for any granule:

$$Word=\{Potency:[(Strong, \mu_{Strong}): P_{Strong}] \ OR/AND \ [(Weak, \mu_{Weak}): P_{Weak}]; \ Evaluation:[(Good, \mu_{Good}): P_{Good}] \ OR/AND \ [(Bad, \mu_{Bad}): P_{Bad}]; \ Activity:[(Active, \mu_{Active}): P_{Active}] \ OR/AND \ [(Passive, \mu_{Passive}): P_{Passive}]\} \quad (5)$$

Where, two antonymic granules, positive and negative, correspond to any criterion. Altogether, there are three universal criteria: Potency, Evaluation, and Activity.

For any of them:

$$P_{Granule}+P_{AntGranule}=1 \quad (6)$$

Positive and negative antonymic granules constitute a sample space: the probability of being strong is opposite to the probability of being weak.

The probability of any granule is equal to the sum of probabilities of its term-units:

$$P_{Granule}= = P_{Granule}(x_1)+ P_{Granule}(x_2)+...+P_{Granule}(x_k) \quad (7)$$

$$P_{AntGranule}= =P_{AntGranule}(x_1)+ P_{AntGranule}(x_2)+ ...+P_{AntGranule}(x_l) \quad (8)$$

Where, the number of term-units in both antonymic granules makes n: $n=k+l$, (9)

with k as the number of term-units in the positive granule, and l as that in negative one.

Then for any term-unit x_i either from positive or negative granule, the probability is:

$$P_{Granule}(x_i)=P_{AntGranule}(x_i)= 1/n \quad (10)$$

The membership function of the positive granule, following the disjunction in fuzzy logic, constitutes:

$$\mu_{Granule} = =max(\mu_{Granule}(x_1), \mu_{Granule}(x_2), ... \mu_{Granule}(x_k)) \quad (11)$$

Where, $\mu_{Granule}(x_1), \mu_{Granule}(x_2), ... \mu_{Granule}(x_k)$ are the membership functions of linguistic term-units of the positive granule.

The same for the negative granule:

$$\mu_{AntGranule}= = max(\mu_{AntGranule}(x_1), \mu_{AntGranule}(x_2), ... \mu_{AntGranule}(x_l)) \quad (12)$$

Where, $\mu_{AntGranule}(x_1), \mu_{AntGranule}(x_2), ... \mu_{AntGranule}(x_l)$ are the membership functions of linguistic term-units of the negative granule.

Thus we can create six differentiating granules for any collocate of the map, narrowing its evaluation to traditional

Osgood's coordinates, but implicitly contributing to them both denotative and connotative estimates of the criteria, to balance the situation when: "Osgood's semantic differential was an attempt to represent words as entities in an n-dimensional space, where measuring the distance between them could naturally follow from our knowledge of Euclidean geometry. Unfortunately, after extensive experimentation, Osgood found his system to rely on "connotative emotions" attached to a word rather than its "denotative meaning", and discontinued further research." [13] Our evaluation absorbs the extended psycholinguistic analysis, since according to J. Deese, associations reveal the latent mental lexicon, the hidden word's value [14, 17].

The example of fuzzy granule-differentiation for some Scottish lexis (He, Kid, and Thow) is notified in Figure 7.

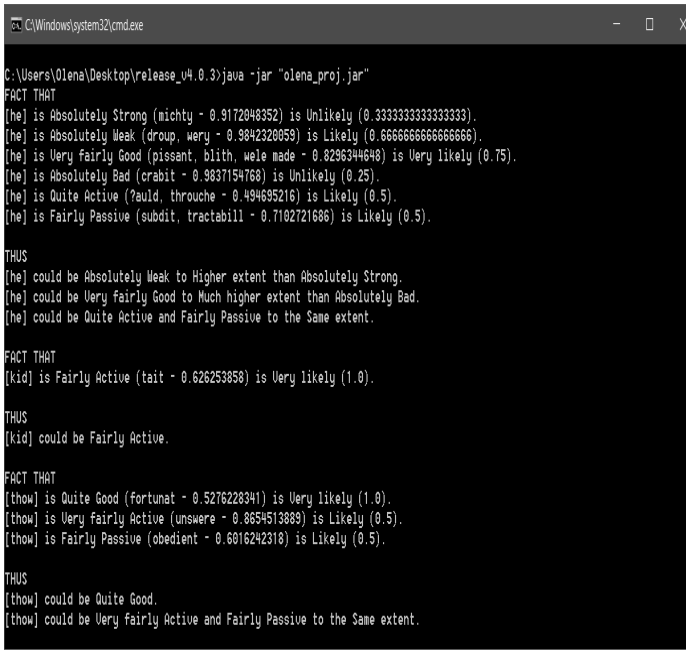


Figure 7: Console with the Older Scottish mental representations for He, Kid, and Thow.

The software reveals how Strong, Weak, Active, Passive, Good, and Bad He, Kid and Thow (in English 'thaw') are on the given map fragment:

Fact that

He is Absolutely Strong (michty - 0.92) is Unlikely (0.33)

He is Absolutely Weak (droup, wery - 0.98) is Likely (0.66)

He is Very Fairly Good (pissant, blith, wele made - 0.83) is Very Likely (0.75)

He is Absolutely Bad (crabit - 0.98) is Unlikely (0.25)

He is Quite Active (ʒauld, throuche 0.49) is Likely (0.5)

He is Fairly Passive (subdit, tractabill - 0.71) is Likely (0.5)

Thus, the software summarizes:

He could be Absolutely Weak to Higher Extent than Absolutely Strong

He could be Very Fairly Good to Much Higher Extent than Absolutely Bad

He could be Quite Active and Fairly Passive to the Same Extent

Thus the richest piece of the mental portrait on early and medieval *Him* has been represented by our fuzzy associative differential.

Kid and Thow are not so amply characterized on the fragment but still well-determined. According to the differential, "The fact, that *kid* is fairly active is very likely". As to *thaw*, the differential says that "Thow could be quite good" and "Thow could be very fairly active and fairly passive to the same extent."

The proposed by us electronic tool, called a fuzzy associative differential, seems to be very beneficial in the (re-)constructing of mental representations, and is highly recommended to the attention of cognitivists.

5.5. Fuzzy Lexis Stratification

On the other hand, the vertical cuts of the associative map in Figure 1 represent stratification figures of Older Scottish lexis [3, 4]. According to the late concept of Lotfi Zadeh, the special way of fuzzy classification was proposed as the way of organizing dictionary, encyclopedias and notebooks. It reveals the special relevance for the historical time-oriented thesaurus like ours.

The fuzzy stratification could open possibilities of ascertaining the lexis stratum gravity poles, time antonymy and significance status of the strata, all enabled by the manipulation with border membership functions [3, 4].

Any stratum in Figure 1 counts 100 years, and there are four of them within our figures for time-structuring of historical vocabulary of Potency, Evaluation, and Activity themes. The proposed by us fuzzy variable:

$$Gravity = \\ = (PositiveConcept/\ \mu(x_i);\ NegativeConcept/\ \mu(x_j)) \quad (13)$$

is introduced in any stratum, where $\mu(x_i)$ and $\mu(x_j)$ are the membership functions of the border (between-strata) antonymic terms x_i and x_j, respectively. The adopted error (the difference in advent time of the border words) makes 50 years.

The stratification for Potency was represented by authors in [3, 4].

Strong-Weak:
1400, (Burely/0.92) – 1420, (Smal/0.86);
1500, (Rik/0.59) – 1500, (Frail/0.61);
1600, (Proud/0.25) – 1600, (Bauch/0.29).
1300-1400 stratum: Gravity = (strong/0.92; weak/0.86);
1400-1500 stratum: Gravity = (strong/0.59; weak/0.61);
1500-1600 stratum: Gravity = (strong/0.26; weak/0.29).

In the same manner we characterize the sectional time antonyms, gravity or concept predominance within the remaining two structures:

Good-Bad:
1400, (Nobil/0.77) – 1400, (Pur/0.92);
1500, (Bony/0.51) – 1500, (Dolly/0.59);
1600, (Braw/0.25) – 1597, (Subaltern/0.28).

1300-1400 stratum: Gravity = (good/0.77; bad/0.92);
1400-1500 stratum: Gravity = (good/0.51; bad/0.59);
1500-1600 stratum: Gravity = (good/0.25; bad/0.28).

Active-Passive:

1420, (Unswere/0.87) – 1400, (Mild/0.93);
1500, (Lifly/0.63) – 1500, (Appliable/0.57);
1602, (Ȝauld/0.32) – 1609, (Submiss/0.17).
1300-1400 stratum: Gravity = (active/0.87; passive/0.93);
1400-1500 stratum: Gravity = (active/0.63; passive/0.57);
1500-1600 stratum: Gravity = (active/0.32; passive/0.17).

The balancing of values or the small predominance of negative concepts (the negative concepts being more alarming are traditionally more developed in equal development conditions) is observed in all the cases, except the latest stratum, which reveals the considerable gravity to the Active flank, pointing out to the notable development of the concept Active throughout the 16-th century. Linguists could speculate on historical reasons, for example, influence of Renaissance ideas in the case. According to Sapir, the language changes remain behind the cultural ones evoking them [28]. Therefore, approximately in the 14-15 centuries, at the time of High Renaissance, the concept Active became comparably more demanded.

5.6. Software Description

The architecture is the same as in [4] and based on MVP design template (*Model-View-Presenter*), delivering the visual reflection and event-procession behaviour into different classes, namely *View* and *Presenter*. In the case of 2RFAN, two XML files of the model DOM are chosen, for positive and negative runs of the 2RFAN.

The software performs altogether 5 main actions. They can be executed by the following button combinations realized in the given map (Figure 3):

- Left Click: Associative antipodes tracing.
- Left Click + Alt: Fuzzy Associative Word Structure evolving.
- Left Click + Shift: Word Differentiation through six scales of the FAD with final evaluative granularity.
- Left Click + Ctrl: Associative Experiment simulation for a chosen stimulus.
- Right click on two words of the same row: Distance and Similarity calculation for any two associates of the same field.

The provided interface is very sensitive and maximally friendly, making its evaluation in words. The correspondence between verbal estimators and membership functions (MFs) is given in Table 1.

The intervals are adopted from [8]. The proposed by us terms differ depending on words they modify: the first column modifies the degree of belief (modifiers for nouns); the second – all the six criteria as well as words Distant and Similar (modifiers for adjectives). In the proposed practice, the conclusion depends on the closeness of MFs to the centric characteristics, in favour of a more intensive modifier if MFs are equidistant from the two centers.

Table1: Correspondence between numeric and lexical evaluation.

Modifying terms for degrees of belief	Modifying terms for good, bad, active, passive, strong, weak; distant and similar	Range	Center
No	Not at All	0.17	0
Very Low	Just a Very Little	0.00 – 0.33	0.17
Low	Just a Little	0.17 – 0.50	0.33
Medium	Quite	0.33 – 0.67	0.50
High	Fairly	0.50 – 0.83	0.67
Very High	Very Fairly	0.67 – 1.00	0.83
Perfect	Absolutely	0.83	1

The terms for granule-probabilities depend on the probability distribution across two antonymic scales. There are four general terms and they are ranged by us in Table 2.

Table 2: Correspondence between granule-probabilities and terms.

Terms for Granule-Probabilities	Range
Very Unlikely	0.00 – 0.24
Unlikely	0.25– 0.49
Likely	0.5 – 0.74
Very Likely	0.75 – 1

The typical granule-estimating Zadeh's phrases [11] are adopted in our fuzzy associative differential (Chapter 5.4 and Figure 7).

The comparative terms for final conclusions on lexis differentiation depend on the number of shifts between the given above four probability ranges for the compared granules, and are given in Table 3.

Table 3: Comparative estimators for final conclusions on lexis differentiation.

Number of Shifts	Comparative Terms
0	the Same Extent
1	Higher Extent
2	Much Higher Extent
3	the Highest Extent

Thus the lexis and syntax of the phrases in Figures 4-7 are clarified by the aforesaid. The principle of computing with words seems to be utterly justified and profitable in linguistic conclusions like these.

6. Discussion

The electronic resource of the type recommends the method of associative modeling for the older languages, and thus extracts the linguistic knowledge of the remote times. The unfolded fragment of the associative map becomes the canvas for speculations on the Older Scottish consciousness. Many-sided conclusions on lexis development could be drawn from the realized on the map simulations. Associative antipodes' extraction reveals the reachability, coherence and deepness of the

map and, on the other hand, between-field associative communication. The chains count the semantic shifts of associations, notifying the reachability potency by the number of between-field transitions. The stability of the found positive and negative associative environment is characterized by the determined fuzziness values. The simulated associative experiment provides the most complete associative answer to the stimulus, meanwhile being sensitive to the forced turnovers in favour of the higher associativity relations. The constructed associative map makes it possible to derive an associative experiment for any map element, equipping it with paradigmatic (synonymic and antonymic), syntagmatic and epidigmatic associations. The fuzzy metrics' identification gives precise figures on similarity and distance between associates inside the field. The lexis differentiation practice provides the complete assessment analysis of nouns and pronouns of the map. The software groups the found characteristics for the regarded lexis into the fuzzy sets that correspond to the six scales Good, Bad, Active, Passive, Strong and Weak, constituting cross-cultural universals of Osgood's differential. The fuzzy sets are supported with joined probability-values and make the fuzzy granules, introducing the partial meanings of both connotative and denotative nature to the criteria of the scales, and breaking the limits of Osgood's differential. The special significance in this respect is referred to our fuzzy associative differential, designed for the purpose. The fuzzy stratification structures help us to draw the historical and lexical conclusions concerning the strata development. The fuzzy facilities are numerous and make the efficient research device, called an electronic FHATh, organized after the principle of computing with words, providing verbalized characteristics to the requested relations.

7. Conclusions

The fragment of a fuzzy model of associative memory provides complete and well-quantified associative relationship, episodically reflecting the world lingual picture of the older speakers, their communicational habits, lingual preferences, inbuilt attitudes and dispositions. The deepness of associative correlation and coherence of the area constitute the subject for the research conclusion. The prospect is observed in multiplication of the simulated network patterns ready to serve different stereotypes. The final combination of them in the larger network would enrich artificial intelligence with cognitive, behavioristic and evolutional linguistic data and mechanisms of their mining, which should broaden its historical prognostication abilities.

References

[1] The Dictionary of the Scots Language, Edinburgh: Scottish Language Dictionaries, 2007. https://www.dsl.ac.uk

[2] O. Basalkevych, "Fuzzy Modelling of Associative maps" IV International Scientific Practical Conference "Linguocognitive and Sociocultural Aspects of Communication", Scientific records of "Ostroh Academy" National University, Ostroh, Ukraine, 56, May 2015.

[3] O. Basalkevych, O. Basalkevych, "Fuzzy Reconstructions in Linguistics" in Proceedings of IEEE Second International Conference on Data Stream Mining and Processing, Lviv, Ukraine, 21-25 August 2018. DOI: 10.1109/DSMP.2018.8478451

[4] O. Basalkevych, "Fuzzy Diachronic Differential" in Proceedings of the IV Ukrainian-German conference "Informatics. Culture. Technics.", Odesa, 106-108, 2016.

[5] J.M.Mendel, "Computing with words: Zadeh, Turning, Popper and Occam" IEEE Computational Intelligence Magazine, 2 (4), 10-17, 2017. https://doi.org/10.1109/MCI.2007.90668978

[6] G.R. Kiss, Ch. Armstrong, R. Milroy, J. Piper, "An associative thesaurus of English and its computer analysis" In Aitken, A.J., Bailey, R.W. and Hamilton-Smith, N. (Eds.), The Computer and Literary Studies, University Press: Edinburgh, 153-165, 1973.

[7] G. Zipf, "Prehistoric 'cultural strata' in evolution of German: the case of Gothic" Modern Language Notes, 62 (8), 522-530, 1947.

[8] L. Martinez, R.M. Rodriguez, F. Herrera, The 2-tuple Linguistic Model. Computing with Words in Decision Making, Springer International Publishing, Switzerland, 2015.

[9] М. Сявавко, Інформаційна Система «Нечіткий експерт», Л.: Видавництво ЛНУім.Франка, 2007.

[10] L. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—I" Elsevier Information Sciences, 8 (3), 1975. https://doi.org/10.1016/0020-0255(75)90036-5

[11] L. Zadeh, "Fuzzy sets and information granularity" Advances in Fuzzy Systems-Applications and Theory Vol 6. Fuzzy Sets, Fuzzy Logic and Fuzzy Systems. Selected Papers by Lotfi A. Zadeh, G.J. Klir and B. Juan (eds.) 433-448. Singapore: World Scientific, 1996.

[12] L. Zadeh, Stratification, Target Set Reachability and Incremental Enlargement Principle. [Electronic Video Resource] The Design of Robotics and Embedded systems, Analysis and Modeling Seminar, 250 Saturdja Dai Hall, UC Berkeley, 2016. https://www.youtube.com/watch?v=Ok0vJ0uykCA&list=PL8092C965854FA5BB&index=30

[13] O. Budanitsky, "Lexical semantic relatedness and its application in natural language processing" Technical report CSRG-390. Computer system research group, 5, 1999.

[14] О.О. Селіванова, Сучасна Лінгвістика. Напрями та Проблеми, Полтава: Довкілля-К, 2008.

[15] A. Jain, D.K. Lobiyal, "Fuzzy Hindi WordNet and Word Sense Disambiguation Using Fuzzy Graph Connectivity Measures" ACM Trans. Asian & Low-Resource Lang. Inf. Process, 15(2): 8:1-8:31, Dec. 2015. https://dx.doi.org/10.1145/2790079

[16] W. Banks, "Linguistic variables: clear thinking with fuzzy logic" Byte Craft Limited, Waterloo (Ontario, Canada), 2013.

[17] M.L. Murphy, Semantic Relations and the Lexicon, Cambridge University Press, 2003

[18] P. Sussner, M. Valle, "Fuzzy associative memories and their relationship to mathematical morphology" In the project of Fuzzy and lattice-based associative memory models, 2012.

[19] U. Priss, "A classification of associative and formal concepts", Napier University Press, 2002.

[20] Oxford Learner's Thesaurus. A Dictionary of Synonyms. Oxford University Press, 2008.

[21] WordNet. A lexical database for the English language. Princeton University, 2006. http:// wordnet.princeton.edu

[22] J. Schafer, Documentation in the O.E.D.: Shakespeare and Nashe as Test Cases, Oxford: Clarendon Press, 1980.

[23] M. Bilynskyi, "Synonymous string as a diachronic reconstruction: the OED Middle English Evidence for verbs and deverbal coinages", Inozemna Philologia, Lviv, 121, 2009.

[24] Historical Thesaurus of the Oxford English Dictionary, Oxford English Dictionary, Oxford University Press, 2019. — www.oed.com

[25] О.В. Бісікало, Концептуальні Основи Моделювання Образного Мислення Людини, Вінниця, 2018.

[26] D. Nelson, C. McEvoy, S. Dennis, "What is and what does free association measure? Free Association" 3-46, 2002.

[27] A. Mochida, "Antonymy and synonymy in adjectives in L2 mental lexicon: Associative or conceptual relations?" The University of Queensland, 2003. https:// www.geocities.jp/akiramochida33/researchpro.html

[28] E. Sapir, "Language and environment" American Anthropologist. New Series, 14 (2), 226-242, 1912.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 234-242 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Transformation of Solar Energy to Electricity with Photovoltaic Systems, Reduction of Electrical Consumption and Installation of LED Lamps in the ITSAO

Jeremias Bravo Tapia*, Juan José Rivera Toledo, Francisco Ramos Guzmán, Yoxkin Estévez Martínez

Instituto Tecnológico Superior de Acatlán de Osorio, División de la Carrera de Ingeniería Electrónica, Carretera Acatlán-San Juan Ixcaquixtla Km. 5.5 Unidad Tecnológica C. P. 74949, Acatlán de Osorio, Puebla, México.

ARTICLE INFO

Article history:
Received: 16 June, 2019
Accepted: 26 August, 2019
Online: 08 October, 2019

Keywords:
Electricity consumption
LED lamps
Photovoltaic system
Solar planter

ABSTRACT

In this work of technological development and research is presented the work done in the ITSAO to contribute to the use of clean energy such as photovoltaics and the implementation of activities to reduce electricity consumption by the technological community, the monitoring of electricity consumption is made on a monthly basis and an annual report is prepared in which the consumption monitoring and the reduction percentage thereof are shown by tables and graphs. The students of electronic Engineering carry out activities of optoelectronics and sustainable development where they intervene with the design and installation of photovoltaic systems that are part of their academic formation. It has a photovoltaic system that produces 2.16 kW-h / d and another will be installed to have photovoltaic energy in the tables of the ITSAO planters; these actions are part of the operational control related to the energy of the Integral Management System (SGI).

1. Introduction.

This paper is an extension of work originally presented in 2018 XXXI International Summer Meeting on Power and Industrial Applications (RVP-AI) [1] off the IEEE Sección México; in the Instituto Tecnológico Superior de Acatlán de Osorio, Puebla, México; work continues with the development of projects related to the saving of electrical energy with renewable energy applications and the replacement of fluorescent tube lamps by LED's lamps; this is how the Academy of Electronic Engineering has intervened in the installation and replacement of LED's tube lamps.

In [2] it presents a work aimed at reducing the consumption of electrical energy delivered by the conventional electricity network, through the use of photovoltaic solar energy. It is proposed the design of an electronic transfer system that facilitates the use of DC electric power from the solar panel, thus eliminating the use of the voltage inverter, commonly used in this type of applications and which increases the cost of the use of solar energy.

Alternative energies have always been present in the daily life of humanity, although their existence has often been unknown because of the need to put the economic over the environmental benefit. In addition to being free, they are clean and protective of the environment [2]. It is necessary to implement an adequate ecological culture in the present generations so that it contributes to the care of the global environment and in this way the new generations are already developed in an environmental system from which it will no longer be necessary to learn to adapt to the environment.

Currently as an example of the application of energy efficiency is in a house at least there is a device that requires the use of electrical energy for its operation, so in some cases there is a high demand and even waste of energy, when there is a considerable amount of these consumer elements, it affects the areas of the rural sector and certain cases of a greater proportion of electricity consumption, because although they have subsidized rates, they suffer increases in costs, thus paying the demand for generated consumption by other sectors, without worrying about the high value in the payment of the energy service. [2].

The study of renewable energies is currently presented as a very useful tool, due to the scarcity of conventional energy sources that have been exploited for several years, without thinking that this would lead humanity to a point where those resources would be exhausted [2]. This allows the development of research for the generation of electricity from clean energy sources in small, medium and large-scale production.

* Jeremias Bravo Tapia., Email: jrmsbt@hotmail.com

Energy efficiency is a primary consideration when designing off-grid renewable energy systems including portable micro-grids. This study focuses on characterizing the potential benefits associated with using energy-efficient exterior area lighting commonly associated with remote installations. Light-Emitting Diode (LED) luminaires are becoming more commercially available, and this study compares two LED products designed for exterior lighting to traditional metal halide lamps. The energy efficiency benefit of the LED luminaires is offset by their lower illuminance. However, a comparison of lighting standards for specific purposes such as security lighting indicates that LEDs may be appropriate for applications where a metal halide system would provide significantly more illumination than required at a much higher energy cost [3].

Measures to improve energy efficiency in small and medium enterprises do not apply due to financial reasons, lack of information and limited internal skills. Suggested information measures, audits, capacity building. As potentially effective measures to support the implementation of energy efficiency measures. Companies in Austria, Bulgaria, Cyprus, Italy, Romania, Slovakia and Spain have identified and implemented a wide range of energy efficiency measures, which in part consist of no-cost options, but are also partly innovative technical solutions [4]. This must be implemented throughout the world starting from our home to be able to impact the environment around us.

Energy efficiency has a vital part to play in all future energy scenarios. Improved energy efficiency limits energy demand growth to one-third by 2040 even though the global economy grows by 150%. Energy policy has traditionally been almost entirely concerned with energy supply and most often focused on the electricity system. The close link between electrification and modernization has given the electricity supply industry in all countries a strong political voice. Energy efficiency, energy conservation or “energy savings” has tended only to be given a high priority in times of crisis, when energy prices are particularly high or when economies face electricity shortages. This was most notable during the “oil crises” of the 1970s. Until recently, energy efficiency policy has always had a lower priority in energy policy development compared to energy supply [5].

All the energy used is ultimately to meet human needs. Natural gas is not burned to heat the room in which one lives by simply heating, but it is done to feel comfortable in that room. Energy is used to produce the materials that are needed to make consumer goods, for example, steel is produced to make cars, paper for printing books and fertilizers to grow more food so as not to be hungry. Energy is not a necessity in itself, it is one of the goods used to meet human needs that can be realized through the use of energy called energy services. [6].

Energy consumption in homes and buildings is large and, in the future, it tends to be unsustainable for our planet, both due to the depletion of natural resources and the irreversible damage it causes to the ecosystem. To help the sustainable development of Humanity it is essential to create an energy education that allows, without waste, continue using fossil fuels for the development of new energy technologies more efficient and in harmony with the environment. The option of renewable energies offers an alternative to achieve a social development that tends to the ecological and that in the future an economic benefit is obtained as well. The energy obtained from photovoltaic panels, specifically the Sun Irradiance, is proposed as an aid to minimize energy consumption in both houses and buildings. Currently, photovoltaic energy is still not profitable, but it seems to be the most viable option until now to solve the great problem of generating electricity without using fossil fuels [7].

The photovoltaic parks have taken great importance as renewable energy generation systems in individual buildings integrating from the design of the building involving the architecture and installers of photovoltaic systems; thus, the photovoltaic industry is a technology that is already included in the architectural design process [8], [9] & [10].

In ITSAO, a photovoltaic system will also be implemented in the gardens of the esplanade so that the technological community has the possibility of connecting their electronic equipment such as laptops and mobile devices outdoors on the rest tables.

2. Determination of solar radiation.

For the development of the photovoltaic part there is location data of the ITSAO in Google Maps [11]. Data provided by NASA corresponding to the solar radiation received by the solar panels, the hours of solar radiation were used to calculate the design of the battery bank and the temperature to see the performance of the solar panels [12].

The coordinates of the location of the ITSAO at a latitude of 18.226883 and a longitude of -98.040143 are shown in Figure 1 a). Figure 1 b) shows the location coordinates on the NASA platform in Prediction of Worldwide Energy Resource Prepare.

Next, Figure 2 shows the data that must be identified for the calculation and design of the photovoltaic system as the minimum radiation for solar panels inclined towards Ecuador, the days without sun in a month to determine the size of the battery bank and the maximum and minimum temperature at two meters high.

Figure 3 a) shows the behavior of the monthly days without solar radiation to add the days of autonomy in the design of the size of the battery bank, in Figure 3 b) the maximum temperature is shown and in Figure 3 c) the minimum temperature is shown at two meters high. this to determine the efficiency behavior of solar panels that depend on the temperature.

One of the very important parameters for the sizing of photovoltaic systems is the solar radiation that is received in the place of the installation of the solar panels, for this data is taken provided by NASA that as mentioned above are reliable data to work sizing, performance, days of autonomy and storage capacity.

Figure 4 shows parameters the solar radiation values in the ITSAO; days without sun, maximum temperature, minimum temperature; in blue are the radiation data for different inclinations are monthly data and an annual average.

Figure 5 shows the optimal peak solar radiation hours (HSP) to calculate the size of the sizing of the photovoltaic system on a monthly and annual basis; The annual average is 5.87 HSP with this the daily production that a solar panel will have is calculated according to its power provided by the manufacturer.

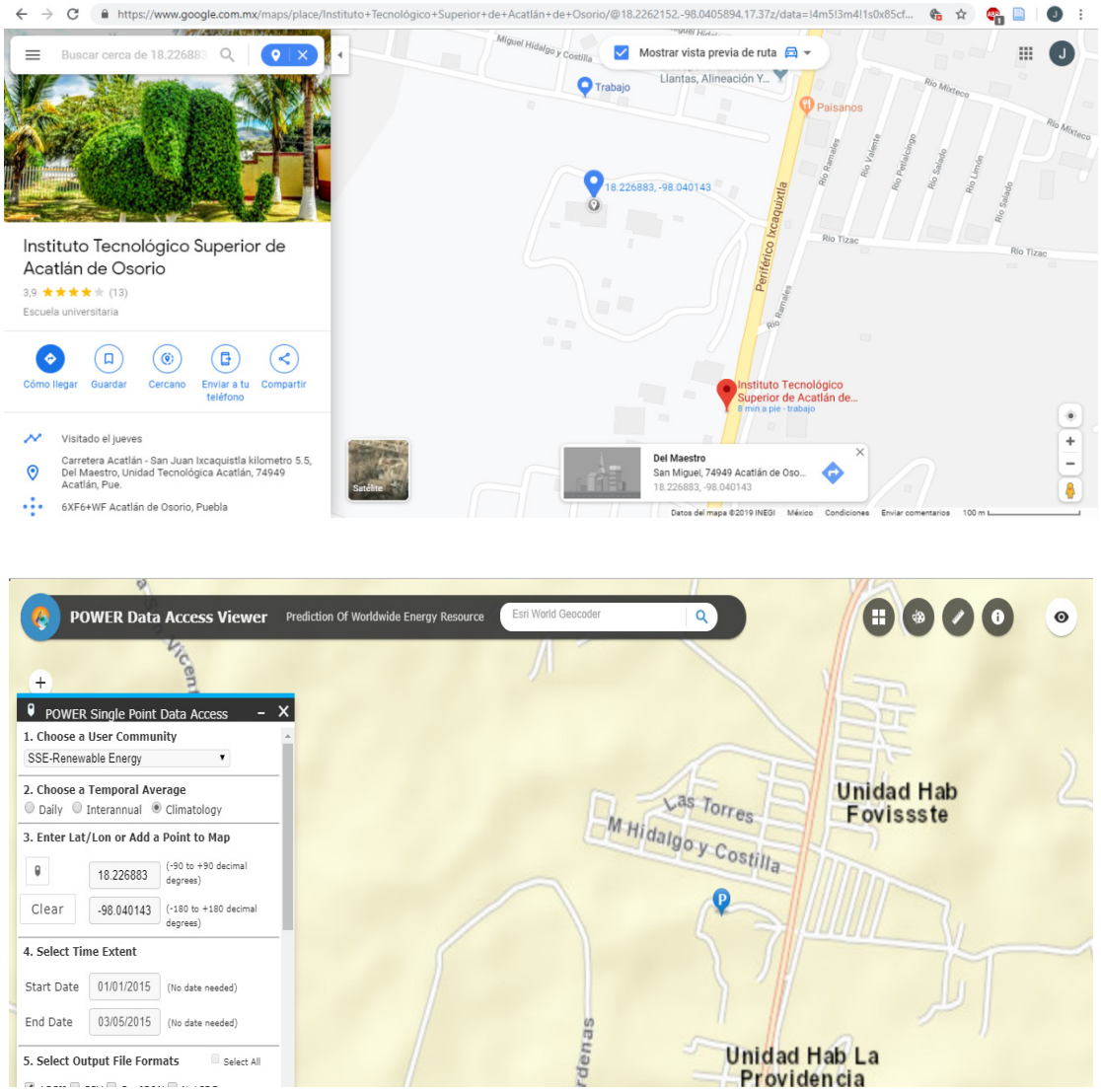
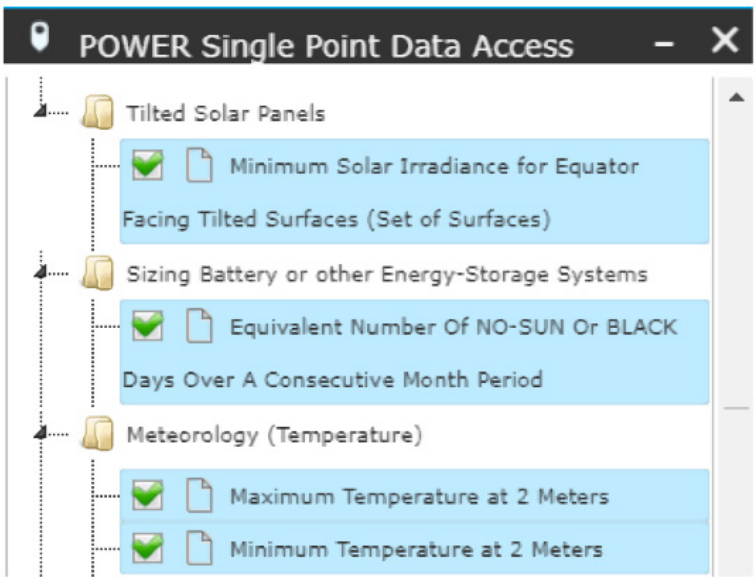
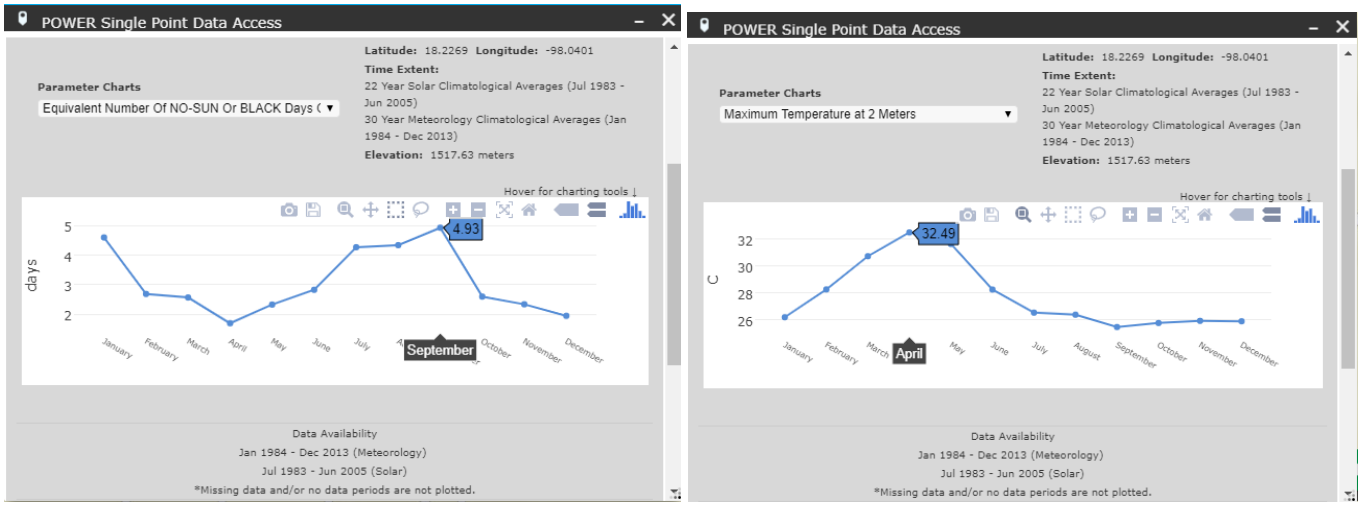


Figure 1: Location coordinates of ITSAO. a) Google maps [11]. b) NASA platform [12].



[12].



b)

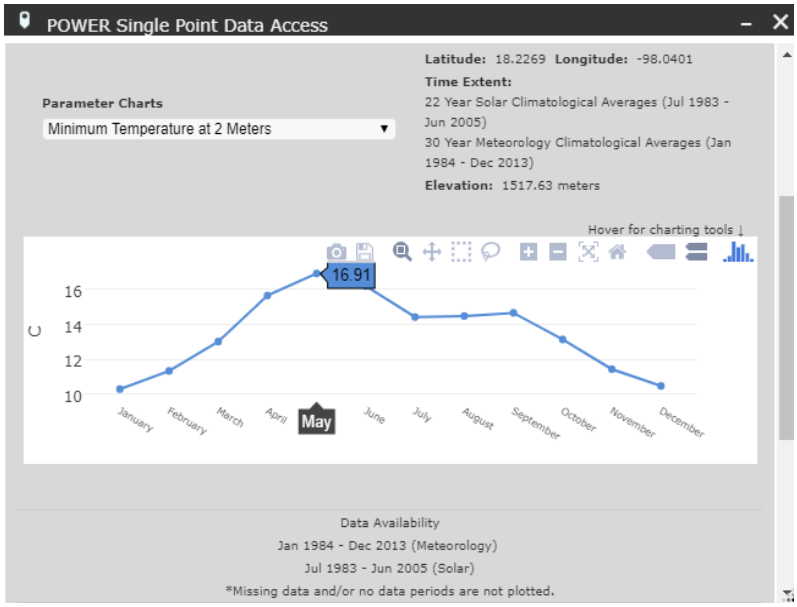


Figure 3: Parameter graphs for calculating photovoltaic systems, a) days without solar radiation, b) maximum temperature and c) minimum temperature [12].

https://power.larc.nasa.gov/downloads/POWER_SinglePoint_Climatology_018d23N_98d40W_15095925.txt

```
-BEGIN HEADER-
NASA/POWER SRB/FLASHFlux/MERRA2/ 0.5 x 0.5 Degree Climatologies
22-year Additional Solar Parameter Monthly & Annual Climatologies (July 1983 - June 2005), 30-year Meteorological and Solar Monthly & Annual Climatologies (January 1984 - December 2013)
Location: Latitude  18.2269   Longitude -98.0401
Elevation from MERRA-2: Average for 1/2x1/2 degree lat/lon region = 1517.63 meters   Site = na
Climate zone: na (reference Briggs et al: http://www.energycodes.gov)
Value for missing model data cannot be computed or out of model availability range: -999
Parameter(s):
EQVLNT_NO_SUN_BLACKDAYS_MONTH    SRB/FLASHFlux 1/2x1/2 Equivalent Number Of NO-SUN Or BLACK Days Over A Consecutive Month Period (days)
T2M_MAX                          MERRA2 1/2x1/2 Maximum Temperature at 2 Meters (C)
T2M_MIN                          MERRA2 1/2x1/2 Minimum Temperature at 2 Meters (C)
SI_EF_MIN_TILTED_SURFACE         SRB/FLASHFlux 1/2x1/2 Minimum Solar Irradiance for Equator Facing Tilted Surfaces (Set of Surfaces) (kW-hr/m^2/day)
SI_EF_MIN_OPTIMAL                SRB/FLASHFlux 1/2x1/2 Minimum Solar Irradiance Optimal (kW-hr/m^2/day)
SI_EF_MIN_OPTIMAL_ANG            SRB/FLASHFlux 1/2x1/2 Minimum Solar Irradiance Optimal Angle (Degrees)
SI_EF_MIN_TILTED_ANG_ORT         SRB/FLASHFlux 1/2x1/2 MinimumSolar Irradiance Tilted Surface Orientation (N/S Orientation)
PARAMETER                JAN     FEB     MAR     APR     MAY     JUN     JUL     AUG     SEP     OCT     NOV     DEC     ANN
-END HEADER-
T2M_MAX                  26.18   28.25   30.72   32.49   31.63   28.24   26.52   26.37   25.45   25.75   25.91   25.88   27.78
T2M_MIN                  10.30   11.34   13.02   15.65   16.91   16.19   14.42   14.48   14.66   13.14   11.44   10.48   13.50
EQVLNT_NO_SUN_BLACKDAYS_MONTH 4.60   2.69    2.57    1.70    2.33    2.83    4.27    4.34    4.93    2.60    2.34    1.95    -999
SI_EF_MIN_TILTED_SURFACE_0   4.37    5.40    6.20    6.50    6.12    5.63    5.66    5.36    4.70    5.08    4.91    4.57    5.38
SI_EF_MIN_TILTED_SURFACE_18  5.14    6.06    6.52    6.39    6.03    5.67    5.66    5.17    4.76    5.52    5.77    5.56    5.69
SI_EF_MIN_TILTED_SURFACE_3   4.53    5.54    6.29    6.52    6.14    5.67    5.69    5.35    4.73    5.18    5.08    4.76    5.46
SI_EF_MIN_TILTED_SURFACE_33  5.48    6.25    6.39    5.92    5.62    5.39    5.35    4.74    4.55    5.56    6.12    6.04    5.62
SI_EF_MIN_TILTED_SURFACE_90  4.01    3.90    3.05    1.90    1.97    2.36    2.20    1.60    2.07    3.17    4.31    4.68    2.94
SI_EF_MIN_OPTIMAL        5.53    6.25    6.52    6.52    6.14    5.70    5.71    5.36    4.78    5.58    6.16    6.17    5.87
SI_EF_MIN_OPTIMAL_ANG    42.00   33.00   20.00   5.00    -6.00   -11.00  -9.00   0.00    12.00   27.00   40.00   46.00   16.00
SI_EF_MIN_TILTED_ANG_ORT  S       S       S       S       N       N       N       S       S       S       S       S       S
```

[12].

https://power.larc.nasa.gov/downloads/POWER_SinglePoint_Climatology_018d23N_98d40W_1e021619.txt

```
-BEGIN HEADER-
NASA/POWER SRB/FLASHFlux/MERRA2/ 0.5 x 0.5 Degree Climatologies
22-year Additional Solar Parameter Monthly & Annual Climatologies (July 1983 - June 2005), 30-year Meteorological and Solar Monthly & Annual Climatologies (January 1984 - December 2013)
Location: Latitude  18.2269   Longitude -98.0401
Elevation from MERRA-2: Average for 1/2x1/2 degree lat/lon region = 1517.63 meters   Site = na
Climate zone: na (reference Briggs et al: http://www.energycodes.gov)
Value for missing model data cannot be computed or out of model availability range: -999
Parameter(s):
EQVLNT_NO_SUN_BLACKDAYS_MONTH    SRB/FLASHFlux 1/2x1/2 Equivalent Number Of NO-SUN Or BLACK Days Over A Consecutive Month Period (days)
T2M_MAX                          MERRA2 1/2x1/2 Maximum Temperature at 2 Meters (C)
T2M_MIN                          MERRA2 1/2x1/2 Minimum Temperature at 2 Meters (C)
SI_EF_MIN_TILTED_SURFACE         SRB/FLASHFlux 1/2x1/2 Minimum Solar Irradiance for Equator Facing Tilted Surfaces (Set of Surfaces) (kW-hr/m^2/day)
SI_EF_MIN_OPTIMAL                SRB/FLASHFlux 1/2x1/2 Minimum Solar Irradiance Optimal (kW-hr/m^2/day)
SI_EF_MIN_OPTIMAL_ANG            SRB/FLASHFlux 1/2x1/2 Minimum Solar Irradiance Optimal Angle (Degrees)
SI_EF_MIN_TILTED_ANG_ORT         SRB/FLASHFlux 1/2x1/2 MinimumSolar Irradiance Tilted Surface Orientation (N/S Orientation)
PARAMETER                 JAN      FEB      MAR      APR      MAY      JUN      JUL      AUG      SEP      OCT      NOV      DEC      ANN
-END HEADER-
T2M_MAX                   26.18    28.25    30.72    32.49    31.63    28.24    26.52    26.37    25.45    25.75    25.91    25.88    27.78
T2M_MIN                   10.30    11.34    13.02    15.65    16.91    16.19    14.42    14.48    14.66    13.14    11.44    10.48    13.50
EQVLNT_NO_SUN_BLACKDAYS_MONTH 4.60  2.69     2.57     1.70     2.33     2.83     4.27     4.34     4.93     2.60     2.34     1.95     -999
SI_EF_MIN_TILTED_SURFACE_0    4.37  5.40     6.20     6.50     6.12     5.63     5.66     5.36     4.70     5.08     4.91     4.57     5.38
SI_EF_MIN_TILTED_SURFACE_18   5.14  6.06     6.52     6.39     6.03     5.67     5.66     5.17     4.76     5.52     5.77     5.56     5.69
SI_EF_MIN_TILTED_SURFACE_3    4.53  5.54     6.29     6.52     6.14     5.67     5.69     5.35     4.73     5.18     5.08     4.76     5.46
SI_EF_MIN_TILTED_SURFACE_33   5.48  6.25     6.39     5.92     5.62     5.39     5.35     4.74     4.55     5.56     6.12     6.04     5.62
SI_EF_MIN_TILTED_SURFACE_90   4.01  3.90     3.05     1.90     1.97     2.36     2.20     1.60     2.07     3.17     4.31     4.68     2.94
SI_EF_MIN_OPTIMAL             5.53  6.25     6.52     6.52     6.14     5.70     5.71     5.36     4.78     5.58     6.16     6.17     5.87
SI_EF_MIN_OPTIMAL_ANG         42.00 33.00    20.00    5.00     -6.00    -11.00   -9.00    0.00     12.00    27.00    40.00    46.00    16.00
SI_EF_MIN_TILTED_ANG_ORT      S        S        S        S        N        N        N        S        S        S        S        S        S
```

[12].

It is possible to determine that there is an annual average of 5.87 hours of minimum solar radiation equivalent to 5.87 kWh/m^2/day. For the photovoltaic calculations applying the rounding, 6 hours of solar radiation are considered.

3. A photovoltaic system installed with LED lamps.

It has a photovoltaic system installed with three 120 W solar panels giving 360 W-h between the three with a daily output of 2.16 kWh/d taking into account 6 hrs of solar radiation for this region. Figure 6 shows the module installed with the three solar panels.

Figure 6: Photovoltaic module installed in the ITSAO.

Figure 7: Battery bank.

Figure 8: Charge controller, batteries and inverter.

The LED lamp has a consumption of 30 W with a reflector installed at an approximate height of 10 m; By having 2.16 kWh/d of production, up to 7 lamps can be connected to work all night, considering a continuous operation of 10 hrs per day with the consumption of 300 W per night of each lamp. Figure 9 shows the installed lamp.

Figure 9: 30 W lamp with reflector installed.

The operational control of energy has the purpose of establishing guidelines and activities for the control of electrical energy in the ITSAO facilities applicable to all activities, processes, products and services that consume electricity.

To reduce electricity consumption to date, we have worked with the acquisition of LED lamp tubes to replace the fluorescent tubes that consume the most electrical energy; in all areas of the ITSAO the lamps have been replaced with the participation of the technological community. Figure 10 shows the LED lamps corresponding to the interior lighting of the ITSAO buildings.

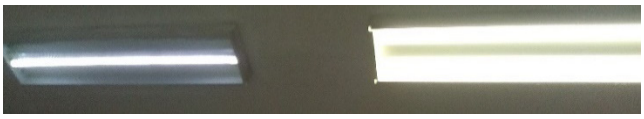


Figure 10: Lamps of LED tubes in operation.

Monthly monitoring of electricity consumption is made by comparing quantitatively the monthly consumption of the current year with the previous one so that the energy savings obtained through an annual report to the representative of the general management can be identified and analyzed annually.

Table 1 shows the monitoring of electricity consumption, in this case, corresponds to the annual report of consumption monitoring in 2016, Table 2 shows the monitoring of electricity consumption for 2017, Table 3 shows the monitoring of electricity consumption for the year 2018 and Table 4 shows the monitoring of electricity consumption for the year 2019.

Table 1: Monitoring of electricity consumption in 2016.

MONITORING OF THE ITSAO ELECTRICAL CONSUMPTION													
YEAR: 2016													
PERIOD	12/29/2014 TO 01/27/2015	01/27/2015 TO 02/26/2015	02/26/2015 TO 03/27/2015	03/27/2015 TO 04/28/2015	04/28/2015 TO 05/28/2015	05/28/2015 TO 06/26/2015	06/26/2015 TO 06/27/2015	06/27/2015 TO 08/27/2015	08/27/2015 TO 09/25/2015	09/25/2015 TO 10/26/2015	10/26/2015 TO 11/26/2015	11/26/2015 TO 12/28/2015	ANNUAL SAVINGS
	12/29/2015 TO 01/27/2016	01/27/2016 TO 02/26/2016	02/26/2016 TO 03/27/2016	03/27/2016 TO 04/27/2016	04/27/2016 TO 05/26/2016	05/26/2016 TO 06/27/2016	06/27/2016 TO 07/25/2016	07/25/2016 TO 08/25/2016	08/25/2016 TO 09/26/2016	09/26/2016 TO 10/25/2016	10/25/2016 TO 11/25/2016	11/25/2016 TO 12/26/2016	
MONTHLY CONSUMPTION kW-h	8960	12180	14000	12040	16800	5180	21280	14840	17640	14000	16100	10640	163660
	8260	11060	9800	11620	18340	13720	12320	15540	16240	19640	15400	10360	162300
SAVINGS IN % WAIT ANNUAL	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	
* SAVINGS IN %	7.81%	9.20%	30.00%	3.49%	-9.17%	-164.9%	42.11%	-4.72%	7.94%	40.29%	4.35%	2.63%	0.83%

Table 2: Monitoring of electricity consumption in 2017.

MONITORING OF THE ITSAO ELECTRICAL CONSUMPTION													
YEAR: 2017													
PERIOD	12/29/2015 TO 01/27/2016	01/27/2016 TO 02/26/2016	02/26/2016 TO 03/27/2016	03/27/2016 TO 04/27/2016	04/27/2016 TO 05/26/2016	05/26/2016 TO 06/27/2016	06/27/2016 TO 07/25/2016	07/25/2016 TO 08/25/2016	08/25/2016 TO 09/26/2016	09/26/2016 TO 10/25/2016	10/25/2016 TO 11/25/2016	11/25/2016 TO 12/26/2016	ANNUAL SAVINGS
	12/29/2016 TO 01/23/2017	01/23/2017 TO 02/23/2017	02/23/2017 TO 03/27/2017	03/27/2017 TO 04/26/2017	04/26/2017 TO 05/25/2017	05/25/2017 TO 06/26/2017	06/26/2017 TO 07/26/2017	07/26/2017 TO 08/24/2017	08/24/2017 TO 08/25/2017	08/25/2017 TO 10/25/2017	10/25/2017 TO 11/24/2017	11/24/2017 TO 12/26/2017	
MONTHLY CONSUMPTION kW-h	8260	11060	9800	11620	18340	13720	12320	15540	16240	19640	15400	10360	162300
	7420	14280	17920	15120	22540	16520	16520	15120	15400	13440	12880	9520	176680
SAVINGS IN % WAIT ANNUAL	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	
* SAVINGS IN %	10.17%	-29.11%	-82.86%	-30.12%	-22.90%	-20.41%	-34.09%	2.70%	5.17%	31.57%	16.36%	8.11%	-8.86%

Table 3: Monitoring of electricity consumption in 2018.

MONITORING OF THE ITSAO ELECTRICAL CONSUMPTION													
YEAR: 2018													
PERIOD	12/29/2016 TO 01/23/2017	01/23/2017 TO 02/23/2017	02/23/2017 TO 03/27/2017	03/27/2017 TO 04/26/2017	04/26/2017 TO 05/25/2017	05/25/2017 TO 06/26/2017	06/26/2017 TO 07/26/2017	07/26/2017 TO 08/24/2017	08/24/2017 TO 08/25/2017	08/25/2017 TO 10/25/2017	10/25/2017 TO 11/24/2017	11/24/2017 TO 12/26/2017	ANNUAL SAVINGS

	12/26/2017 TO 01/25/2018	01/25/2018 TO 02/23/2018	03/23/2018 TO 03/26/2018	03/26/2018 TO04/24/2018	04/24/2018 TO 05/25/2018	05/25/2018 TO 06/25/2018	06/25/2018 TO 07/26/2018	07/26/2018 TO 08/24/2018	08/24/2018 TO 09/25/2018	09/25/2018 TO 10/25/2018	10/25/2018 TO 11/26/2018	11/26/2018 TO 12/26/2018	
MONTHLY CONSUMPTION kW-h	7420	14280	17920	15120	22540	16520	16520	15120	15400	13440	12880	9520	176680
	6720	10360	16940	13160	17080	18340	16520	15400	20720	19460	14000	10640	179340
SAVINGS IN % WAIT ANNUAL	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	
* SAVINGS IN %	9.43%	27.45%	5.47%	12.96%	24.22%	-11.02%	0.00%	-1.85%	-34.55%	-44.79%	-8.70%	-11.76%	-1.91%

Table 4: Monitoring of electricity consumption in 2019.

MONITORING OF THE ITSAO ELECTRICAL CONSUMPTION													
YEAR: 2019													
PERIOD	12/26/2017 TO 01/25/2018	01/25/2018 TO 02/23/2018	03/23/2018 TO 03/26/2018	03/26/2018 TO04/24/2018	04/24/2018 TO 05/25/2018	05/25/2018 TO 06/25/2018	06/25/2018 TO 07/26/2018	07/26/2018 TO 08/24/2018	08/24/2018 TO 09/25/2018	09/25/2018 TO 10/25/2018	10/25/2018 TO 11/26/2018	11/26/2018 TO 12/26/2018	ANNUAL SAVINGS
	12/26/2018 TO 01/26/2019	01/26/2019 TO 02/22/2019	02/22/2019 TO 03/25/2019	03/25/2019 TO 04/24/2019	04/24/2019 TO 05/27/2019	05/27/2019 TO 06/24/2019	06/24/2019 TO 06/25/2019	06/25/2019 TO 08/24/2019	08/24/2019 TO 08/25/2019	08/25/2019 TO 10/25/2019	10/25/2019 TO 11/26/2019	11/26/2019 TO 12/26/2019	
MONTHLY CONSUMPTION kW-h	6720	10360	16940	13160	17080	18340	16520	15400	20720	19460	14000	10640	179340
	7140	9940	20300	15960	22120	17920	20720						114100
SAVINGS IN % WAIT ANNUAL	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	
* SAVINGS IN %	-6.25%	4.05%	-19.83%	-21.28%	-29.51%	2.29%	-25.42%	100.00%	100.00%	100.00%	100.00%	100.00%	36.38%

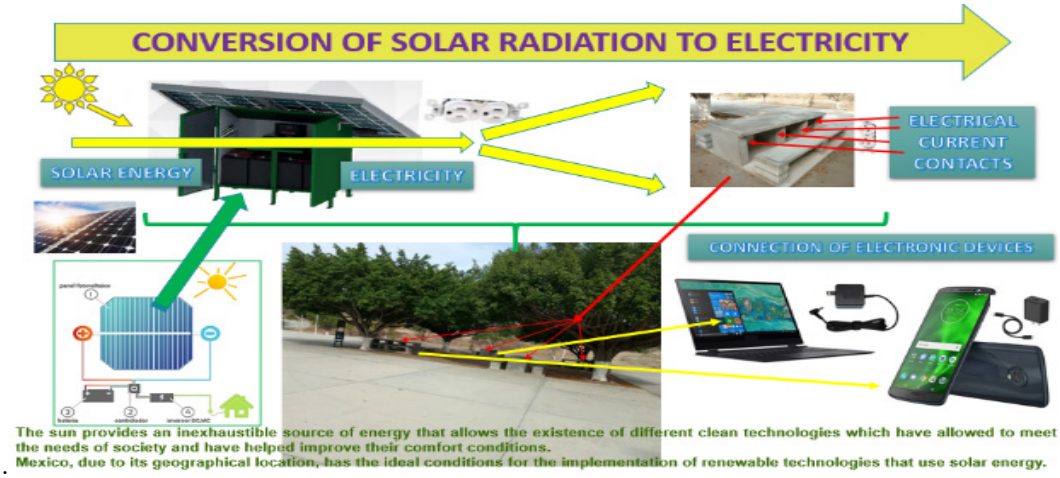


Figure 11: Infographic of the photovoltaic system for the ITSAO planters.

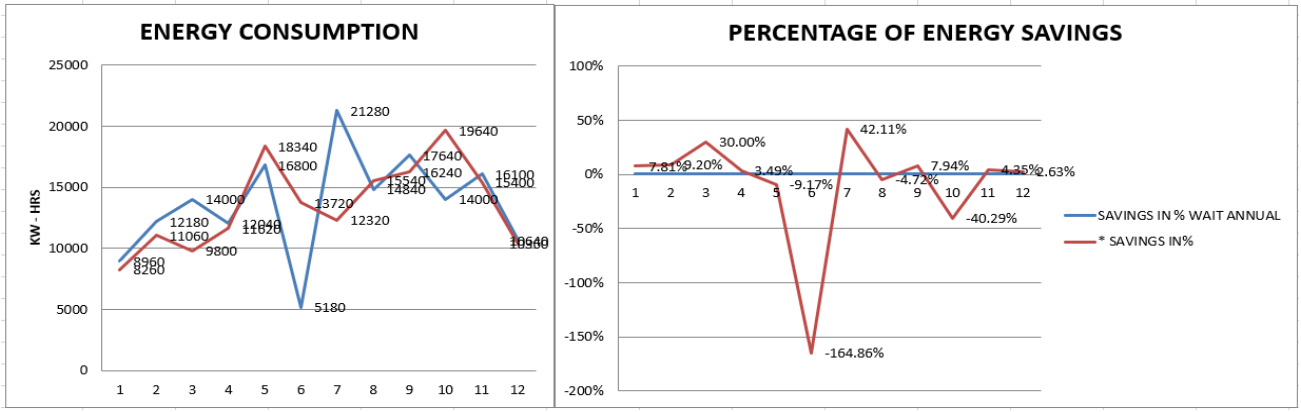


Figure 12: Graph of consumption in 2016 compared to 2015 and Graph of the reduction percentual for electricity consumption in 2016 compared to 2015.

One of the activities to be implemented is a photovoltaic system with the purpose of supplying electricity to the tables located in the planters of the esplanade; the conversion of solar energy to electricity will be applied so that the technology community can use it to power their electronic devices such as laptops and mobile devices while they rest or gather in the flower beds. Figure 11 shows an infographic that describes the operation and location of the photovoltaic elements and electrical power outlets

5. Results of the electric follow-up corresponding to the reports of 2016, 2017, 2018 and the current follow-up in 2019.

Figure 12 shows the comparison charts of electricity consumption and shows the relative comparison of electricity consumption reduction for 2016 compared to 2015.

Figure 13 shows the comparison charts of electricity consumption and shows the relative comparison of electricity consumption reduction for 2017 compared to 2016. In this case, electricity consumption increased due to the installation of air conditioning equipment in salons that did not have this electrical equipment.

Figure 14 shows the comparison charts of electricity consumption and shows the relative comparison of electricity consumption reduction for 2018 compared to 2017. In this case, electricity consumption increased due to the use of air conditioning equipment in salons that did not have this electrical equipment.

Figure 15 shows the comparison charts of electricity consumption and shows the relative comparison of electricity consumption reduction for 2019 compared to 2018. In this case, to date only monitoring is carried out until July the follow-up is in process.

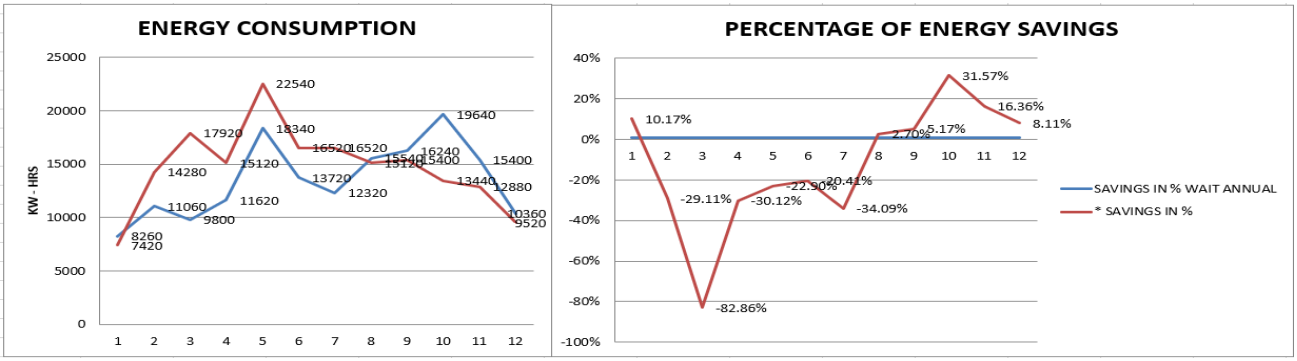


Figure 13: Graph of consumption in 2017 compared to 2016 and Graph of the reduction percentual for electricity consumption in 2017 compared to 2016.

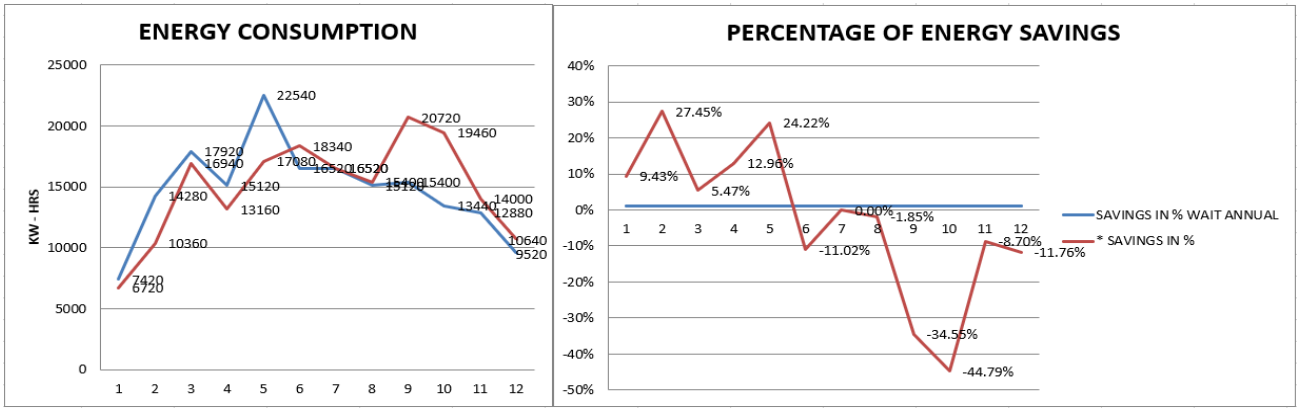


Figure 14: Graph of consumption in 2018 compared to 2017 and Graph of the reduction percentual for electricity consumption in 2018 compared to 2017.

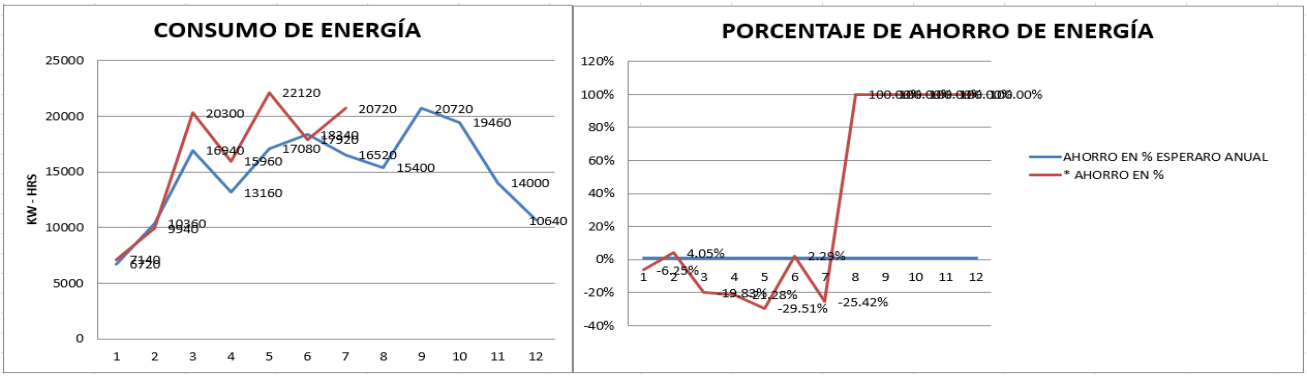


Figure 15: Graph of consumption in 2019 compared to 2018 and Graph of the reduction percentual for electricity consumption in 2019 compared to 2018.

The activities implemented are not only part of the Integrated Management System of the ITSAO, but they are also part of the academic and professional training of the students and the intervention of the Electronic Engineering Academy. They are also part of the research area corresponding to the research line "Renewable Energy and Energy Efficiency".

6. Conclusions.

With the execution of the activities carried out, the emphasis is mainly on the use of clean energies such as solar as an alternative means to obtain electricity for the external lighting of the ITSAO facilities, with the implementation of LED lamps it is possible to reduce the electricity consumption with sustainable approach; the entire technological community will be given the opportunity to use photovoltaic energy for their outdoor electronic devices. The students manage to carry out sustainable development practices, research foundations, research workshops and optoelectronics to strengthen their academic training. The activities implemented to reflect a good result by observing the graphs in the reduction of electricity consumption.

Conflict of Interest.

The authors declare no conflict of interest.

Acknowledgement.

The research was carried out with the support of the Electronic Engineering area and the ITSAO technological community within the framework of the implementation of the Environmental Management System and the research area.

References.

[1] B. T. Jeremias, E. M. Yoxkin, J. R. T. Juan, y R. G. Francisco, "Photovoltaic Energy, LED Technology and Energy Efficiency in the ITSAO", en *2018 XXXI INTERNATIONAL SUMMER MEETING ON POWER AND INDUSTRIAL APPLICATIONS (RVP-AI)*, 2018, pp. 19-22.

[2] J. C. Cruz-Ardila, J. C. Cardona-Gómez, y D. M. Hernández-Porras, "Aplicación electrónica para el ahorro de energía eléctrica utilizando una energía alternativa", *Entramado*, vol. 9, n.º 2, pp. 234-248, jul. 2013.

[3] A. N. Sperber, A. C. Elmore, M. L. Crow, y J. D. Cawlfield, "Performance evaluation of energy efficient lighting associated with renewable energy applications", *Renew. Energy*, vol. 44, pp. 423-430, ago. 2012.

[4] J. Fresner, F. Morea, C. Krenn, J. Aranda Uson, y F. Tomasi, "Energy efficiency in small and medium enterprises: Lessons learned from 280 energy audits across Europe", *J. Clean. Prod.*, vol. 142, pp. 1650-1660, ene. 2017.

[5] Steven Fawkes, Kit Oung, y David Thorpe, *Best Practices and Case Studies for Industrial Energy Efficiency Improvement. An Introduction for Policy Makers*. DTU Library, 2016.

[6] J. de Beer, *Potential for Industrial Energy-Efficiency Improvement in the Long Term*. Springer Netherlands, 2000.

[7] Soraya Navarro Rayas, José Antonio González Moreno, y César M. López Andrade, "Implementación de un sistema fotovoltaico para la alimentación de un edificio de usos múltiples". 2016.

[8] S. Medio, "Photovoltaic Design Integration at Battery Park City, New York", *Buildings*, vol. 3, n.º 2, pp. 341-356, abr. 2013.

[9] M. Wichliński y J. Staszkiewicz, "Photovoltaic installation for a roofed car park in Egypt as an example of effective use of solar energy", *E3S Web Conf.*, vol. 49, p. 00127, 2018.

[10] C.-G. Haba, "Monitoring photovoltaic parks for damage prevention and optimal operation", en *2017 International Conference on Electromechanical and Power Systems (SIELMEN)*, Iasi, 2017, pp. 321-326.

[11] "Google Maps", *Google Maps*. [Online]. Available in: https://www.google.com.mx/maps/place/Instituto+Tecnol%C3%B3gico+Superior+de+Acatl%C3%A1n+de+Osorio/@18.2248389,-98.0409622,17z/data=!3m1!4b1!4m5!3m4!1s0x85cf4182fb37cc4f:0xf1a9691f53f1acf!8m2!3d18.2248389!4d-98.0387735. [Accessed: Mar 03, 2019].

[12] "POWER Data Access Viewer". [Online]. Available in: https://power.larc.nasa.gov/data-access-viewer/. [Accessed: Mar 03, 2019].

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 243-249 (2019)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Improve the Accuracy of Short-Term Forecasting Algorithms by Standardized Load Profile and Support Regression Vector: Case study Vietnam

Nguyen Tuan Dung[*,1], Nguyen Thanh Phuong[2]

[1]*Ho Chi Minh Power Company (EVNHCMC), Vietnam*

[2]*Ho Chi Minh Ho Chi Minh University of Technology (Hutech), Vietnam*

ARTICLE INFO

Article history:
Received: 13 June, 2019
Accepted: 02 September, 2019
Online: 08 October, 2019

Keywords:
Short-term load forecast
Regression model
Standardized Load Profile
Support Vector Regression

ABSTRACT

Short-term load forecasting (STLF) plays an important role in building business strategies, ensuring reliability and safe operation for any electrical system. There are many different methods, including: regression models, time series, neural networks, expert systems, fuzzy logic, machine learning and statistical algorithms used for short-term forecasts. However, the practical requirement is how to minimize the forecast errors to prevent power shortages or wastage in the electricity market and limit risks.
The paper proposes a method of short-term load forecasting by constructing a Standardized Load Profile (SLP) based on the past electrical load data, combining machine learning algorithms Support Regression Vector (SVR) to improve the accuracy of short-term forecasting algorithms.

1. Introduction

Load forecasting is a topic of electrical systems which has been studied for a long time. There are two main approaches in this area: Traditional statistical modeling of the relationship between load and factors affecting load (such as time series, regression analysis, etc) and artificial intelligence, machine learning methods. Statistical methods assume load data according to a sample and try to forecast the value of future loads using different time series analysis techniques. Intelligent systems are derived from mathematical expressions of human behavior / experience. Especially since the early 1990s, neural networks have been considered one of the most commonly used techniques in the field of electrical load forecasting, because it assumes that there is a nonlinear function related to historical values and some external variables with future values may affect the output [1]. The approximate ability of neural networks has made their applications popular.

In recent years, an intelligent calculation method involving Support Vector Machines has been widely used in the field of load forecasting. In 2001, Bo-Juen Chen, Ming-Wei Chang, and Chih-Jen Lin used the Support Vector Regression technique to solve the electrical load prediction problem (forecasting a maximum daily load of the next 31 days). This was a competition organized by EUNITE (European Network on Intelligent Technologies for Smart Adaptive Systems). Information was provided includes: demand data of the past two years, daily temperature of the past four years and local holiday events. Data was divided into 2 parts: a part used for training (about 80 - 90%) and the rest used for algorithm testing (about 20-10%). The set of training inputs included: data of the previous day, the previous hour, the previous week, the average of the previous week. Their approach in fact won the competition [2].

Since then, there have been several studies exploring the different techniques used for optimizing SVR to perform load forecasting [3-10]. The main reason for using SVM in load forecasting is that it can easily model the load curve, the relationship between the load and the dynamics of changing load demand (such as temperature, economic and demographics).

However, there are some problems encountered when the above algorithms apply to reality:

- Climate conditions always play an important role in load forecasting. They show the relationship between climate and load demand, when we do the load forecasting for the post-test period, it is very difficult to forecast the values of weather and climate used as the input of the algorithm and these values are often not available.

[*] Nguyen Tuan Dung., Email: jrmsbt@hotmail.com

- Electrical load samples include hidden elements, which tend to be similar to the previous load model. However, it will lead to a false forecast of the following days if the date pattern is different from the previous day or there is an event that impacts. Therefore, the use of the dataset (training inputs included: data of the previous day, the previous hour, the previous week, the average of the previous week) has many risks if the load models are not identical.
- If the forecast time frame is greater than the past data frame (more than 07 days due to the algorithm data is the previous week's values), there will be a lack of input to run the algorithm.
- In addition, for Asian countries (such as Vietnam) that use lunar calendar, one of the most difficult and unpredictable issues is the Lunar New Year (usually in late January or early February), or the lunar calendar (Hung King's Anniversary), etc. There is a deviation between the solar calendar and the lunar calendar (the load models are not identical). Therefore, it often leads the forecast results of algorithm for this period with large errors.

For this reason, the paper proposes a solution to build a Standardized Load Profile (SLP) based on the historical load dataset as a training dataset. This input dataset is combined with the Support Vector Regression algorithm (SVR) to improve the accuracy of short-term forecast results, solve the problem of deviation between the solar and the lunar calendar, as well as overcome the input data frame.

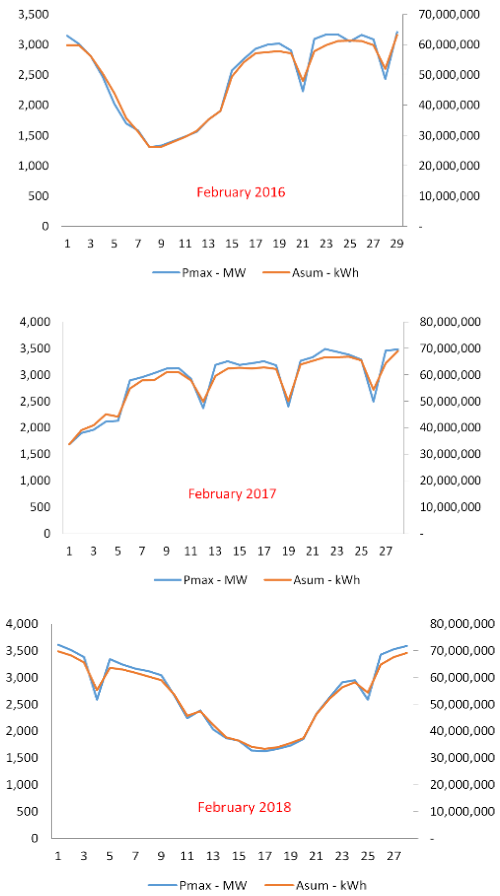


Figure 1: The load profiles of February over the years

SLP will be built for all 365 days and 8,670 cycles in a year. SLP will be an important dataset during training, testing and forecasting process. SLP will be built for all 365 days and 8,670 cycles in 1 year. SLP will be an important set of data during training, testing and forecasting. SLP will standardize load models: by hours, by days, by seasons, and by special day types (including lunar dates). Therefore, SLP will contribute to solve the above-mentioned difficulties and improve the quality of electrical load forecasting.

2. Methodology

Observing the load profiles of February of Ho Chi Minh City over the years (Figure 1), a huge fluctuation in chart shape over the years can be seen. This results in the use of historical data for forecasting this period of time is extremely complicated.

In fact, the algorithms used to forecast in Vietnam have to go through an intermediary which converts these months into regular months (without holidays, Lunar New Year). After being calculated, the forecast result will be reversed or the result will be accepted with a large error. Commercial software provided by foreign countries all have this problem.

2.1. Standardized Load Profiles (SLP)

The Standardized Load Profile is an electrical load profile according to the relative values, converted from the total power consumption during the electrical load research cycle. The standardized load profle of day / month / year of each electrical load sample is constructed by dividing the load profile of a sample (from the measured data collected by day / month / year) by the power consumption of day / month / year of the sample.

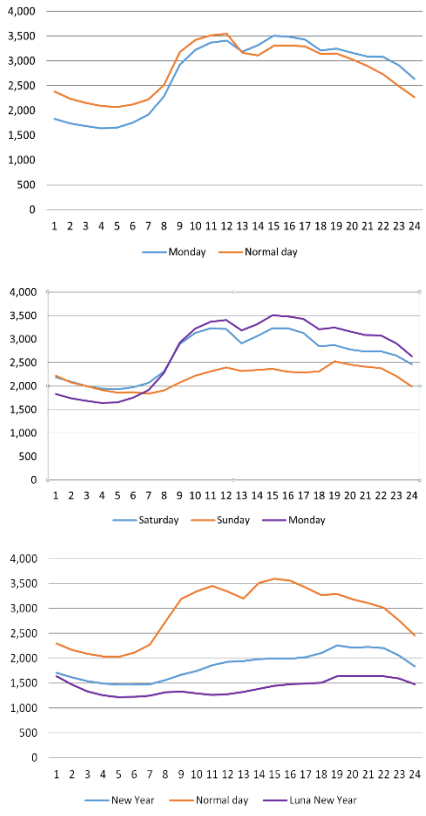


Figure 2: Typical load profiles of some days of a year

Considering the load profiles of the days in a week and some special holidays of the year in Ho Chi Minh City area (Figure 2),the difference between weekdays (from Tuesday to Friday) can be ignored and they have the same load chart. For the load profiles on Monday, they are different from the normal days at 0:00 to 9:00, due to the forwarding demand from Sunday.

For load profiles on Saturday, there is an insignificant change compared to normal days, mainly the load demand decreases in the evening due to the start of the weekends. Particularly for load profiles on Sunday, it is completely different from normal days (the demand for electricity is low).

When observing the load chart of the New Year and Lunar New Year, a noticeable difference can be seen where the graphs are almost flat and the load demand is quite low because these are holidays.Particularly on Lunar New Year, the load demand is the lowest since this is the longest holiday of the year (maybe from 6 to 9 days).

Standardized Load Profiles (SLP) are built by taking the value of the collected capacity in a 60-minute period divided by its maximum capacity. We need to build SLP for 365 days per year. Some typical SLP:

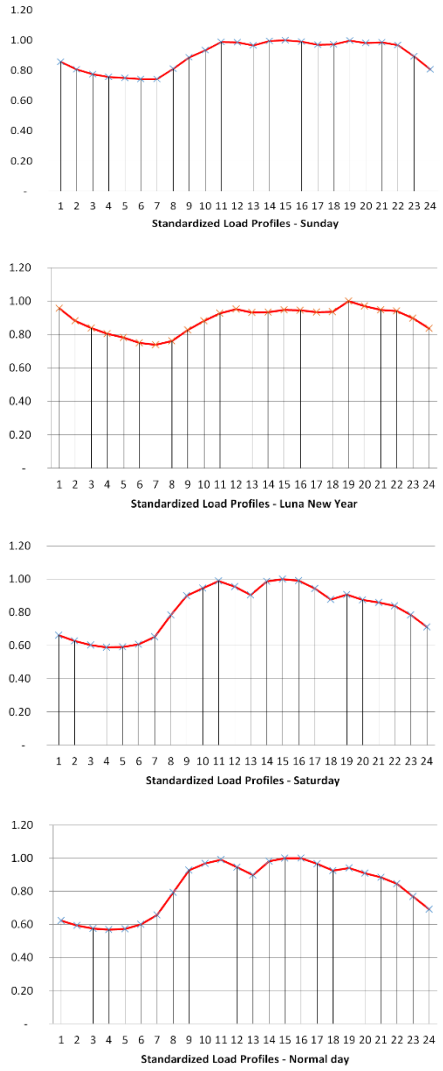


Figure 3: SLP of some days in a year

Based on the SLP of each cycle of the past data set, we can build the SLP data set for future forecast periods. This should be accurate to each cycle, each type of day (holidays, weekdays, working days, holidays, etc.), each week and month.Therefore, the standardized load profiles (SLP) is a special feature and is also an important input parameter of the SVR (NN) machine learning algorithms training process to rebuild the load curves, from which we can estimate the amount of lost or not recorded data during the measurement process.

2.2. Support vector regression (SVR)

The SVM was proposed by Vapnik in [7] to solve the data classification problem. Two years later, the proposed version of SVM was successfully applied to non-linear regression problems. This method is called support vector regression (SVR) and it is the most common form of SVMs.

The goal of SVR is to create a model that predicts unknown outputs based on known inputs. During training, the model is formed based on the known training data set (x_1, y_1), (x_2, y_2), ..., (x_n, y_n), where x_i is input vectors and y_i is output vectors. During the test period, the model was trained on the basis of new inputs $x_1, x_2, ..., x_n$ to make predictions about unspecified outputs $y_1, y_2, ..., y_n$.

Consider a known training set $\{x_k, y_k\}$, k = 1, ..., N with input vectors $x_k \in Rn$ and scalar output vectors $y_k \in R$. The following regression model can be developed by using the nonlinear mapping function φ (.): $R_n \rightarrow R_{nh}$ to map the input space into a multidimensional characteristic space and build linear regression in it, as shown in (1):

$$\psi(\xi) = \overline{T})(\xi) + \beta \tag{1}$$

Where ω represents the weight vector and b is the deviation. The optimization problem is formed in the original space in (2):

$$\min_{\omega,b,\xi,\xi^*} \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \tag{2}$$

that subjects to the constraints shown in (3):

$$\begin{aligned} y_i - (\omega^T\varphi(x_i) + b) &\le \varepsilon + \xi_i, \\ (\omega^T\varphi(x_i) + b) - y_i &\le \varepsilon + \varepsilon_i^*, \end{aligned} \tag{3}$$

$$\xi_i, \xi_i^* \ge 0, i = 1,2,...,N$$

where xi is mapped in a multidimensional vector space with the mapping φ, ξ_i is the upper limit of the training error and ξ_i^* lower. C is the constant that determines the error cost, that is, the tradeoff between the complexity of the model and the accepted larger degree of error. The parameter ε includes the width of the non-sensitive area, which is used to match the training data [7-10]. The parameters C and ε are not known in advance and must be determined by some mathematical algorithm applied on the training set (eg Grid - Search and Cross - Validation). The goal of the SVR is to place many input vectors x i inside the pipe

$\left|y_i - (\omega^T \varphi(x_i) + b)\right| \le \varepsilon$, shown in Figure 4. If the xi is not in the tube, the errors ξi, ξ_i^* will occur.

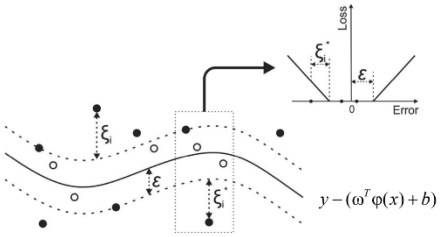


Figure 4: ε tube of nonlinear SVR

To solve the optimization problem identified by (2) and (3), it is necessary to develop a dual problem using Lagrange function, the weight vector ω and the deviation b. The SVM results for the regression model in the double form are shown in (4), where α_i and α_i^* are the Lagrange multipliers, K (x_i, x) represents the Kernel function, defined as a midpoint $\varphi(x_i)^T$ và $\varphi(x_i)$.

$$y(x) = \sum_{i=1}^{N} (\alpha_i^* - \alpha_i) K(x_i, x) + b \qquad (4)$$

The Kernel functions allow the calculation of dot product in a feature space of height using the input data from the original space, without explicit computation φ (x). The Kernel function is often used in non-linear regression problems, which is used in this study, as the radial basis function (RBF) presented in (5):

$$K(x_i, x) = \varphi(x_i)^T \varphi(x) = e^{-\gamma \|x_i - x\|^2}, \gamma > 0 \qquad (5)$$

where γ represents the Kernel parameter, which should also be determined by mathematical algorithms. More information about SVR can be found in [5] – [6], [11] – [12].

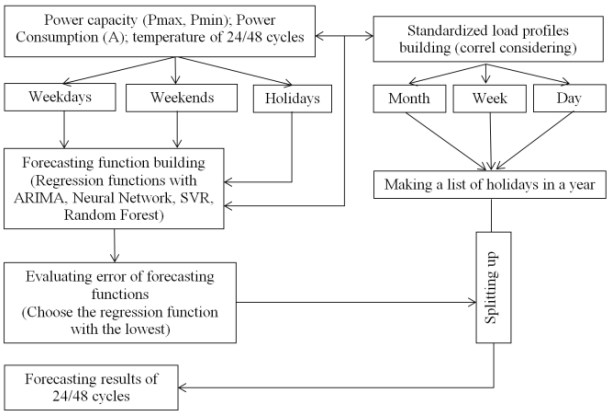


Figure 5: Flowchart of forecasting algorithm by SLP – SVR

2.3. Research models

Processed historical data (power consumption, capacity, temperature recorded at 24 cycles - 60 minutes each) with the Standardized load Profiles (SLP) will be included in modules to build regression functions under SVR (Support Vector Regression), NN (Neural Network) algorithms to build regression functions.

Then we use the above data set to check and evaluate the error of regression functions. After that we choose the regression function with the smallest error which will be used as regression function for the next forecast phase.

The SLP data set in 24 cycles of the expected period (including holidays, etc.) and the forecasted temperature in 24 cycles of the corresponding period will be the input for the regression function that is selected to export forecast results in 24 cycles for a period of 7 - 30 days.

3. Results and Discussion

3.1. Input data:

The article uses data from January 1, 2015, to November 17, 2018, of EVNHCMC to run test models. After pretreatment, the data set is divided into 2 volumes: training set and test set, in which the test set is the last 30 days of the data set. Or the data set is divided into phases to test the forecast results in different time periods.

Input data for training algorithms include: capacity (Pmax/Pmin) in 60-minute cycles; temperature (max / min) in 60-minute cycles; standardized load profiles of 24 hours of day; list of holidays and Lunar New Year in the forecast year.

A useful measurement parameter is the mean absolute percentage error (MAPE) which is used to evaluate the error of models.

$$MAPE = \frac{1}{n} \sum 100 \frac{\left|Y_t - Y_t^f\right|}{Y_t} \qquad (6)$$

The algorithms are programmed in Mathlab language and the results are exported to Excel files for data exploitation.

3.2. SVR Models

Processed historical data (power consumption, capacity, temperature recorded at 24 cycles - 60 minutes each) with the Standardized load Profiles (SLP) will be included in modules to build SVR models, with: normalization coefficient C, width of pipe ε and Kernel function; 4 typical SVR model parameters are proposed:

Table 1: SVR model parameters

Model	C - BoxConstraint	ε - Epsilon	KernelFunction
SVR 1	93.42	32.5	Polynomial
SVR 2	500.32	0.01	Gaussian
SVR 3	1	50.03	Linear
SVR 4	100	0.01	Linear

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Do not use abbreviations in the title or heads unless they are unavoidable.

3.3. RFR models

A set of regression trees with each set of different rules to perform non-linear regression. The algorithm builds a total of 20 trees, with a minimum leaf size of 20. The size of leaves is smaller or equal to the size of the tree to control overfitting and bring about high performance [13] - [14]. The algorithm uses the same input data set of models.

RFR model (Random Forest Regression) is a method of constructing regression models from historical data and is also a machine learning method like current advanced models. Therefore it is used as a result to compare with the proposed SVR model.

3.4. Neural Network Models

We use Feedforward Neural Network models with the input variables and training data set as above. A-hidden-layer network architecture with class size of 10 and Sigmoid activation function is used. At the same time, the usual Neural network with 3-hidden-layer network architecture, in which: the first hidden layer has a size of 10; The second hidden layer has a size of 8 and the third hidden layer has a size of 5.

FNN model (Feedforward Neural Network) is a method of constructing regression models from historical data and is also a machine learning method like current advanced models. Therefore it is used as a result to compare with the proposed SVR model.

4. Results and Analysis

Run the forecast results for February 2018 (the month of the Lunar New Year) to assess the degree of error of the models

4.1. The model with inputs included: data of the previous day, the previous hour, the previous week and the average of the previous week

Processed historical data (power consumption, capacity, temperature recorded at 24 cycles - 60 minutes each) with the Standardized load Profiles (SLP) will be included in modules to build regression functions under SVR,Neural Network and Random Forest algorithms to build regression functions.

Table 2: Results of checking errors of regression models

Date	Ytr	Yts1	Yts2	Yts3	Yts4	YtNN	Ytfeed	YtRF
23/1/18	9.71	4.05	5.02	6.35	4.19	6.09	4.55	2.91
24/1/18	8.30	3.65	2.61	7.00	4.25	0.65	4.76	4.19
25/1/18	7.17	4.35	3.57	7.42	4.21	4.58	5.84	4.63
26/1/18	7.10	6.20	6.77	7.48	6.39	6.58	5.82	6.44
27/1/18	9.22	1.37	0.44	3.27	1.33	0.56	1.91	1.06
28/1/18	9.68	2.16	3.28	7.12	0.32	25.51	5.89	3.93
29/1/18	9.15	5.30	6.17	6.92	4.91	5.71	5.96	5.67

We choose the regression function with the smallest error whichwill be used as regression function for the next forecast phase.The model Yts4 is selected to be a forecasting model.

- Forecast results for February 2018

Considering the forecast results for February of the model, we see a big difference between reality and forecasting. The reason is that we used the historical data of January 2019 (7-14-30 days before the forecasting date) as the input for the training model.

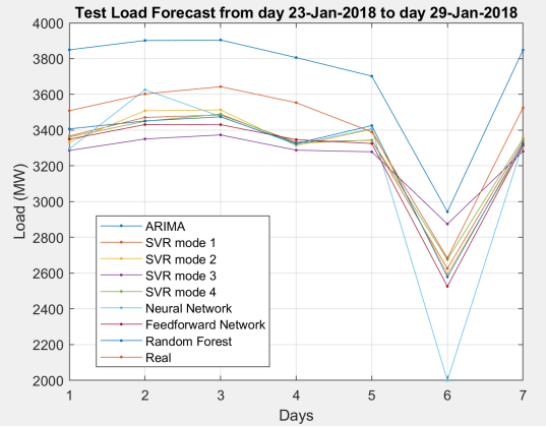


Figure 6: Regression models test

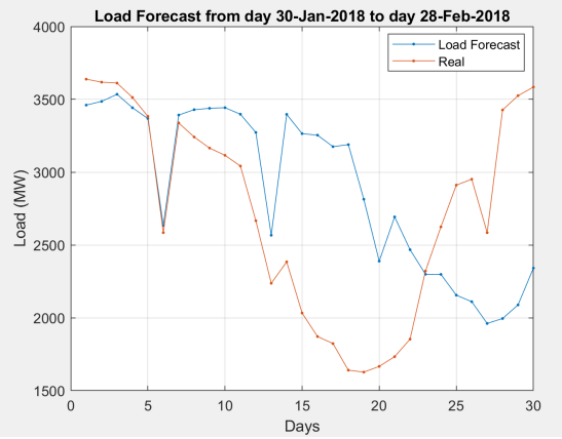


Figure 7: Forecast results for the next 30 days

4.2. SLP - SVR combination model

Processed historical data (power consumption, capacity, temperature recorded at 24 cycles - 60 minutes each) with the Standardized load Profiles (SLP) will be included in modules to build regression functions under SVR, Neural Network and Random Forest algorithms to build regression functions

- Results of testing SVR models

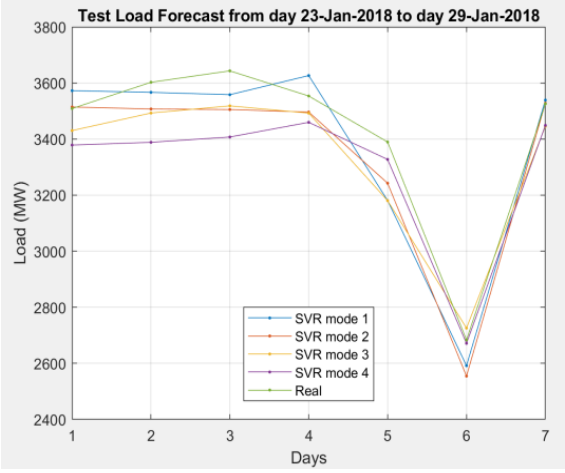


Figure 8: SVR models test

Table 3: Results of checking errors of SVR models

Date	Yts1	Yts2	Yts3	Yts4
23/1/18	1.15	0.64	2.22	3.87
24/1/18	1.70	2.12	2.95	6.19
25/1/18	3.03	3.30	3.38	6.68
26/1/18	1.35	1.04	1.76	2.76
27/1/18	6.77	4.56	6.42	1.56
28/1/18	4.18	5.09	1.81	0.76
29/1/18	0.24	0.12	2.69	2.14
MAPE	2.63	2.41	3.03	3.42

- Results of testing machine learning models

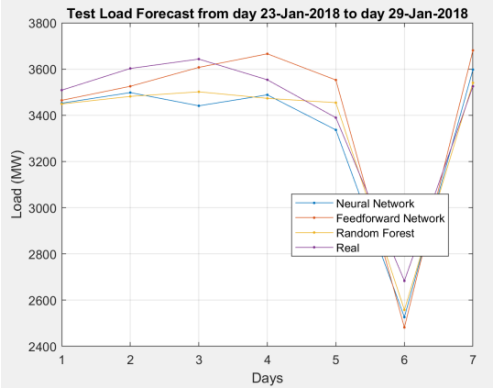


Figure 9: Machine learning models test

Table 4: Results of checking errors of machine learning models

Date	YtNN	YtFeed	YtRF
23/1/18	1.25	1.61	1.70
24/1/18	2.14	2.90	3.36
25/1/18	0.99	5.55	3.89
26/1/18	3.16	1.84	2.26
27/1/18	4.81	1.56	1.92
28/1/18	7.51	5.85	4.68
29/1/18	4.41	2.05	0.43
MAPE	3.47	3.05	2.60

- Results of testing regression models:

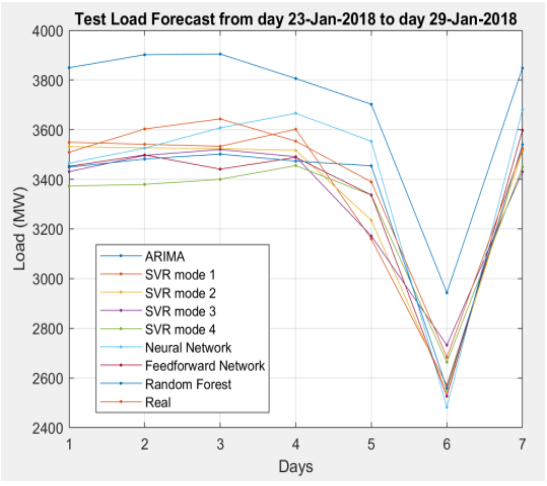


Figure 10: Regression models test

Table 5 - Results of checking errors of all models

Date	Ytr	Yts1	Yts2	Yts3	Yts4	YtNN	Ytfeed	YtRF
23/1/18	9.71	1.15	0.64	2.22	3.87	1.25	1.61	1.70
24/1/18	8.30	1.70	2.12	2.95	6.19	2.14	2.90	3.36
25/1/18	7.17	3.03	3.30	3.38	6.68	0.99	5.55	3.89
26/1/18	7.10	1.35	1.04	1.76	2.76	3.16	1.84	2.26
27/1/18	9.22	6.77	4.56	6.42	1.56	4.81	1.56	1.92
28/1/18	9.68	4.18	5.09	1.81	0.76	7.51	5.85	4.68
29/1/18	9.15	0.24	0.12	2.69	2.14	4.41	2.05	0.43
MAPE	8.62	2.63	2.41	3.03	3.42	3.47	3.05	2.60

The results in this Table 5 is the test run results of the regression models being developed. The evaluation of the MAPE results of the models aims to select a standard model for the official forecasting model in the later stage. Considering models Yts2 (2.41%) and YtRF (2.60%), they all have quite low error results. However, when considering the error according to each component, the model Yts2 has more advantages and the error of each component is also lower than YtRF. Therefore, it is appropriate for the author to choose the model Yts2.

We choose the regression function with the smallest error which will be used as regression function for the next forecast phase.The model Yts2 is selected to be a forecasting model.

- Forecast results for February 2018

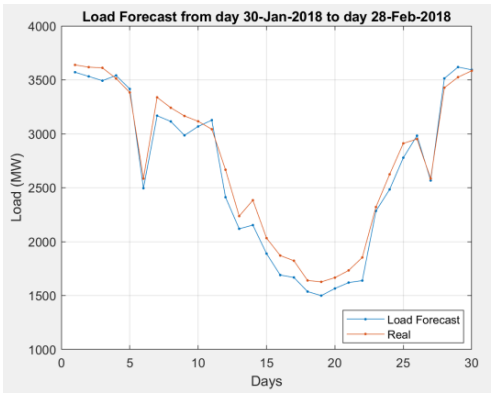


Figure 11: Forecast results for the next 30 days

5. Conclusion

Observe the experimental results in the forms of testing datasets (load data sets of the previous day, the previous week, the previous month and the dataset of Standardized Load Profile - SLP), we see the results of the SLP-SVR model are close to the actual value of February 2018, while the results of the old model are in quite large deviation.

Thus, through experimentation, we see that the use of Standardized Load Profile (SLP) as the input dataset for modules of the forecasting regression function is effective and give forecasting results with low errors. It solves the problem of deviation between the solar and the lunar dates, especially in the months of lunar new year, as well as resolving the difference between the solar and lunar cycles.

References

[1] M H M R ShyamaliDilhani and ChawalitJeenanunt, Daily electric load forecasting: Case of Thailand. 7th International Conference on Information Communication Technology for Embedded Systems 2016 (IC-ICTES 2016). 978-1-5090-2248-9/16/$31.00 ©2016 IEEE.

[2] Juan Huo,Tingting Shi and Jing Chang, Comparison of Random Forest and SVM for Electrical Short-term Load Forecast with Different Data Sources. 978-1-4673-9904-3/16/$31.00 ©2016 IEEE.
[3] Lemuel Clark P. Velasco, Christelle R. Villezas and Jerald Aldin A. Dagaang, Next Day Electric Load Forecasting Using Artificial Neural Networks. 8th IEEE International Conference Humanoid, Nanotechnology, Information Technology Communication and Control, Environment and Management (HNICEM). The Institute of Electrical and Electronics Engineers Inc. (IEEE) – Philippine Section 9-12 December 2015 Water Front Hotel, Cebu, Philippines.
[4] Electricity Load Forecasting for the Australian Market Case Study version 1.3.0.1 by David Willingham. https://www.mathworks.com/matlabcentral.
[5] Nguyen Tuan Dung, Tran Thu Ha, Nguyen Thanh Phuong, Page(s):90 – 95. 10.1109/GTSD.2018.8595514. 2018 4th International Conference on Green Technology and Sustainable Development (GTSD).
[6] E. Ceperic, V. Ceperic, and A. Baric, A strategy for short-term load forecasting by support vector regression machines, IEEE Transactions on Power Systems, vol. 28, pp. 4356–4364, Nov. 2013.
[7] V.Vapnik, 1995, "The nature of statistical learning theory," Springer, NY.
[8] S.R. Gunn, 1998: Support Vector Machines for Classification and Regression, Technical Report, Image Speech and Intelligent Systems Research Group, University of Southampton.
[9] V. Cherkassky, Y. Ma, 2002: Selection of Meta-parameters for Support Vector Regression, International Conference on Artificial Neural Networks, Madrid, Spain, Aug. pp. 687 – 693.
[10] D. Basak, S. Pal, D.C. Patranabis, Oct. 2007: Support Vector Regression, Neural Information Processing – Letters and Reviews, Vol. 11, No. 10, pp. 203 – 224.
[11] A.J. Smola, B. Schölkopf, Aug. 2004: A Tutorial on Support Vector Regression, Statistics and Computing, Vol. 14, No. 3, pp. 199 – 222. 0960-3174 ©2004 Kluwer Academic Publishers.
[12] Understanding Support Vector Machine Regression and Support Vector Machine Regression, http://www.mathworks.com.
[13] Breiman L.: Random Forests. Machine Learning 45 (1), 5-32 (2001). http://dx.doi.org/10.1023/A:1010933404324.
[14] Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. Classification and Regression Trees, Wadsworth, Belmont, CA, 1984. Since 1993 this book has been published by Chapman & Hall, New York.

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 250-257 (2019)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Design and Simulation of a Doppler-Radar RF Front-End Transceiver

Mohamed Elhefnawy*

Department of Electrical Engineering, Faculty of Engineering, October 6 University, 12564, Egypt

ARTICLE INFO

Article history:
Received: 27 July, 2019
Accepted: 20 September, 2019
Online: 08 October, 2019

Keywords:
RF front-end
Microwave components
Electromagnetic simulation tools

ABSTRACT

This paper presents the use of Electromagnetic simulation tools such as ADS and CST in order to practice designing the RF front-end transceiver for a continuous-wave radar. Teaching the microwave engineering course based on the conventional education methods results in students who may not be able to apply their theoretical knowledge to design the microwave components. This work introduces a project-based education which is expected to guide the undergraduate students toward the design and simulation of the most important active and passive microwave components such as oscillator, power amplifier, Cantenna, low noise amplifier, mixer, filter and power divider. This project can be added to the curriculum of the microwave engineering course to achieve a balance between the theoretical and practical knowledge.

1. Introduction

The fundamentals of designing and simulating many microwave components are overviewed in this paper. The microwave engineering course curriculum needs to be modified in order to prepare the undergraduate students to meet the industry requirements [1]. The RF front-end transceiver for a continuous-wave radar has been developed using Advanced Design System (ADS) and Computer Simulation Technology (CST) tools. The RF front-end transceiver is a perfect platform for the undergraduate students to understand the design and simulation of microwave components. This paper presents an overview about the operation and structure of the Doppler radar; then the design and simulation of each microwave component are introduced; finally, the whole RF front-end transceiver for a continuous-wave radar is simulated using ADS.

2. Operation and Structure of a Doppler Radar

The Doppler radar transmits a continuous-wave at a specific frequency to an object; the frequency of the received reflected wave from the moving object is shifted due to Doppler effect. The velocity of the moving object can be calculated based on the relationship between the velocity and the Doppler frequency. Figure 1 shows the block diagram of the proposed RF front-end transceiver for a continuous-wave radar.

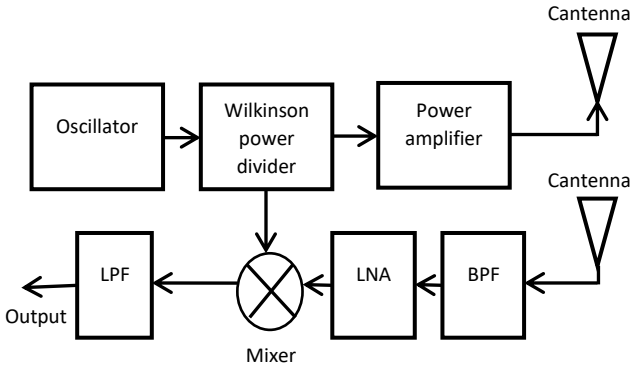


Figure 1: Block diagram of a Doppler-Radar RF Front-End Transceiver

There are two different Cantennas, one of them is used for transmitting the continuous-wave while the reflected wave is received by the other Cantenna. A power amplifier is used to increase the coverage area. The continuous wave is generated by a 10 GHz local oscillator. The Wilkinson power divider is used to equally divide the power of the oscillator between the power amplifier and the mixer. The reflected wave is received by the receiving Cantenna, and then it is filtered by a microstrip coupled line band pass filter (BPF) and amplified by a low noise amplifier (LNA). The output from LNA is applied to the microwave single balanced mixer. The mixer generates the Doppler frequency in addition to the other harmonics. The low pass filter (LPF) is

* Mohamed Elhefnawy., Email: mmmelhefnawy.eng@o6u.edu.eg

implemented to remove the harmonics and pass the Doppler frequency.

3. Design and Simulation of Microwave Components

The microwave components of RF front-end transceiver for a continuous-wave radar are designed and simulated using ADS; also, these microwave components are made from microstrip lines with default substrate properties, except the metal Cantenna which is designed and simulated using CST.

3.1. Design and Simulation of 10 GHz Oscillator

The two-port transistor oscillator consists of the resonant circuit and the active circuit. The active circuit is connected with the resonant circuit to provide the required negative resistance that can cancel out the resistance of the resonant circuit; so that the resonant frequency with no damping can be generated. The design procedures of the proposed two-port oscillator are started by selecting a transistor (ATF-13786) which has a high degree of instability, then performing the stability checks using the output stability circles in ADS to find the unstable region on the Smith chart, where the Smith chart is stable at its center because $S_{11} < 0$ and $S_{22} < 0$. Then the reflection coefficient (Γ_L) is selected from the unstable region on the Smith chart; such that the magnitude of the input reflection coefficient (Γ_{in}) is maximized, where Γ_{in} is obtained as follows

$$\Gamma_{in} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1-S_{22}\Gamma_L} \quad (1)$$

The load impedance (Z_L) can be obtained from the following expression:

$$Z_L = Z_o \frac{1+\Gamma_L}{1-\Gamma_L} \quad (2)$$

The large value of negative resistance at the input port of the active device can be obtained as the magnitude of the input reflection coefficient is increased. The input impedance (Z_{in}) is given by

$$Z_{in} = Z_o \frac{1+\Gamma_{in}}{1-\Gamma_{in}} = R_{in} + jX_{in} \quad (3)$$

The required conditions for getting oscillation are $R_S + R_{in} < 0$ and $X_S + X_{in} < 0$, where R_S and X_S are the source tuning network resistance and reactance, respectively; the source tuning network impedance (Z_S) is selected to be $Z_S = R_S + jX_S = \frac{|R_{in}|}{3} - jX_{in}$. The source reflection coefficient can be calculated by

$$\Gamma_S = \frac{Z_S - Z_o}{Z_S + Z_o} \quad (4)$$

The following expression is used to find the output reflection coefficient

$$\Gamma_{out} = S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1-S_{11}\Gamma_S} \quad (5)$$

Finally, a single stub matching network is designed to convert a 50 Ω load to Z_L ; and the source matching network is designed to convert R_S into Z_S [2]. The complete circuit design for the proposed 10 GHz oscillator is simulated using ADS as shown in Figure 2.

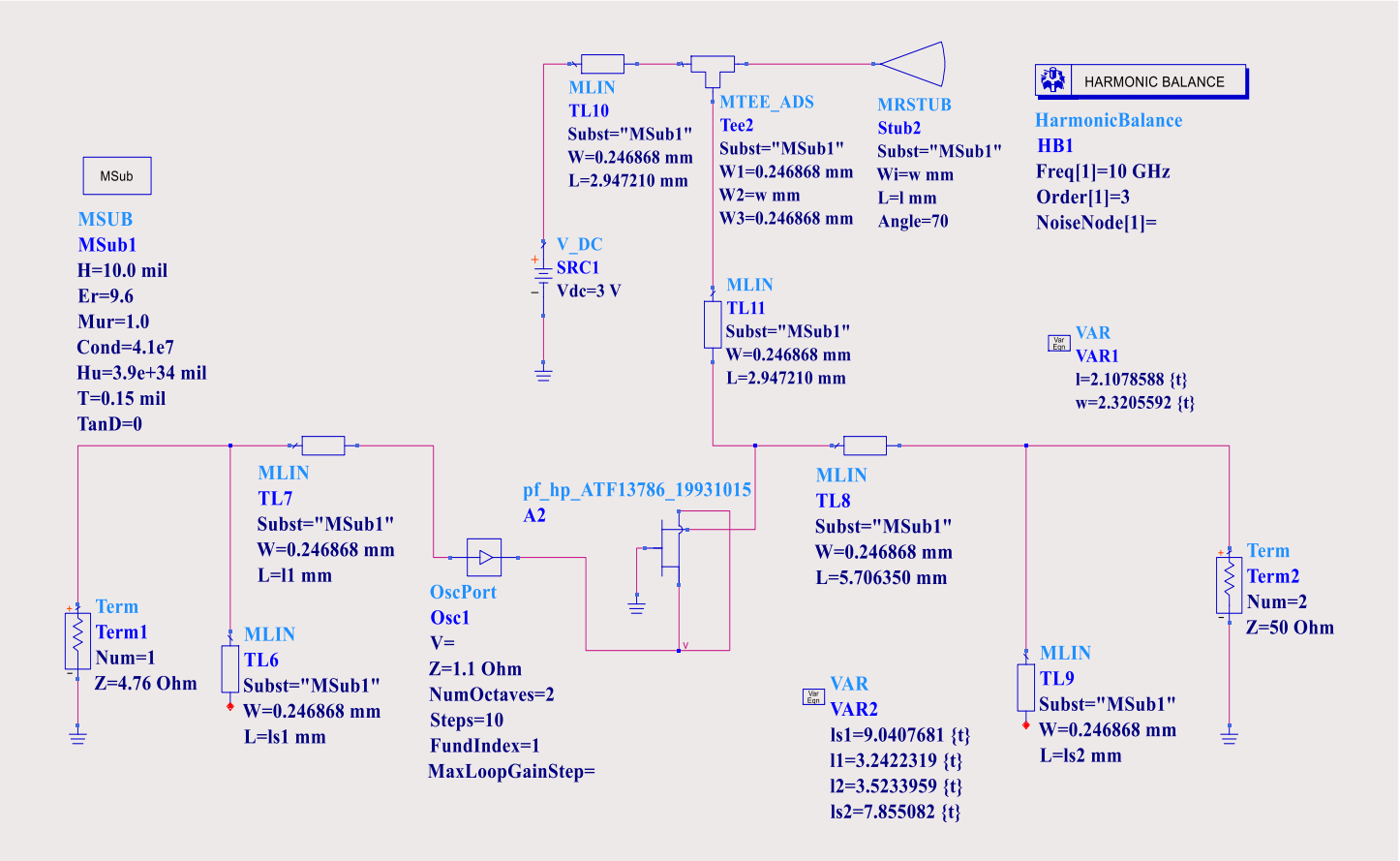


Figure 2: The Complete Circuit Design for the Proposed 10 GHz Common Gate Oscillator

3.2. Design and Simulation of a Wilkinson Power Divider

The Wilkinson power divider is matched at its three inputs. Port 1 receives the power and splits that power equally between ports 2 and 3, where these ports are isolated by using a resistor of 100 Ω; and each of them is connected with port 1 through an arm. In Wilkinson power divider, all ports have 50 Ω impedances; there are two arms each one has a length of a quarter-wavelength and an impedance of 70.71 Ω. Wilkinson power divider is simulated in ADS as shown in Figure 3 [2].

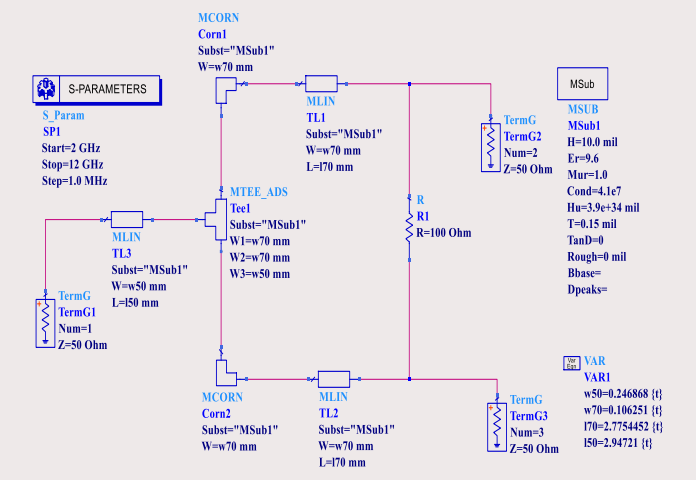


Figure 3: Simulation of Wilkinson Power Divider using ADS

3.3. Design and Simulation of a Class-A Power Amplifier

Class-A power amplifier is used in the final stages of the transmitters to increase the power level at the output of the transmitter. This power amplifier class can provide a minimum distortion because it is always biased on over the entire range of the RF input signal cycle. The main difference between class-A power amplifier and the small-signal amplifier is that the output network is optimized to obtain large voltage and current swing at the same time at the output of the device. The design of the class-A power amplifier starts by the selection of a suitable device. The transistor Cree's CGHV1J025D is chosen because it can operate up to a frequency of 18 GHz. The drain to source voltage (V_{DS}) is selected to be 15 V. The dc analysis using dc FET curve tracer in ADS is used to find the gate to source voltage (V_{GS}) which can provide a drain to source current (I_{DS}) of 858 mA. Then the stability check is performed. The maximum output power and the power added efficiency (PAE) can be plotted on the Smith chart as a function of the load and source impedances by using the load-pull and source-pull simulations templates in ADS under large-signal operation conditions, so that the optimal load and source impedances relevant to the maximum output power and power added efficiency can be determined. The obtained maximum output power is 7.35 W and the PAE is 40.68 %. The input and output matching networks are designed to convert 50 Ω into the obtained optimal source and load impedances, respectively. The complete circuit design for a class-A power amplifier is shown in Figure 4 [3].

3.4. Design and Simulation of a Cantenna

The transmitting and receiving antennas of the proposed radar are selected to be circular waveguide antenna; this antenna type is called Cantenna, it is consisted of a circular waveguide and a monopole soldered on a probe for coaxial-cable to waveguide transition as shown in Figure 5. The total length of the circular waveguide is 0.75 λ_G, where λ_G is the standing wavelength inside the waveguide. The probe is located at a distance of 0.25 λ_G from the closed aperture of the Cantenna. The length of the monopole is 0.25 λ, where λ is the wavelength of the transmitted signal. The radius of the Cantenna can be obtained by $\boldsymbol{radius} = \frac{\chi'_{11}}{\beta_{c,11}}$, where $\boldsymbol{\chi'_{11}}$ is the first root of the first Bessel function derivative for a circular waveguide at TE_{11} mode; and $\boldsymbol{\beta_{c,11}}$ is the cut-off phase constant [4].

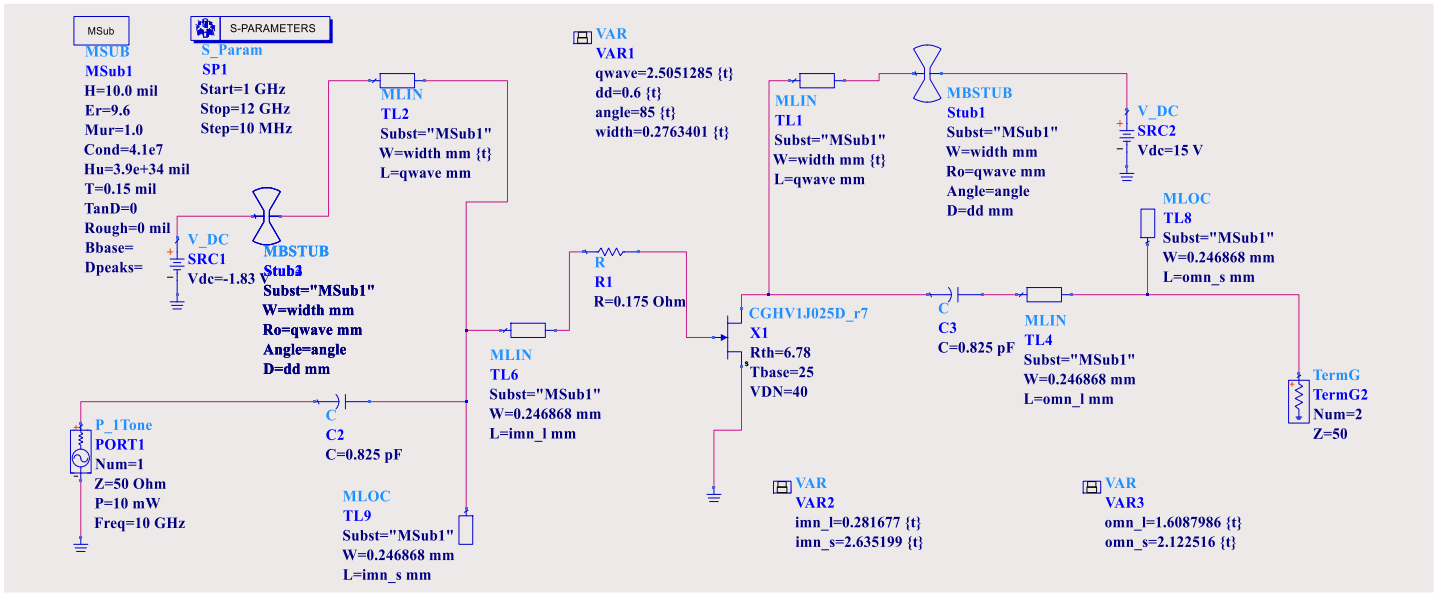


Figure 4: The Complete Circuit Design for the Proposed Class-A Power Amplifier

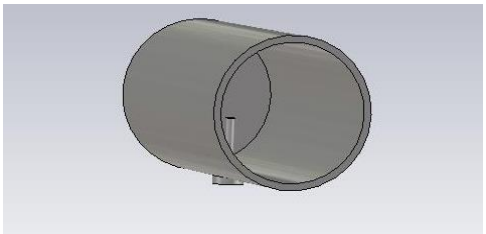


Figure 5: Structure of the Proposed Cantenna

3.5. Design and Simulation of a Microstrip Coupled Line Band-Pass Filter

A microstrip coupled line band-pass filter (BPF) is designed and simulated using ADS as shown in Figure 6. The type of this filter has been selected to be a fifth order Chebyshev filter with 0.5 dB ripple. The admittance inverter for the first coupling section is calculated by using the following equation

$$Z_o J_1 = \sqrt{\frac{\pi\Delta}{2\, g_1}} \tag{6}$$

For the intermediate coupling sections, the following design equations are used to find the admittance inverter for each coupling section:

$$Z_o J_2 = \frac{\pi\Delta}{2\sqrt{g_1 g_2}} \tag{7}$$

$$Z_o J_3 = \frac{\pi\Delta}{2\sqrt{g_2 g_3}} \tag{8}$$

$$Z_o J_4 = \frac{\pi\Delta}{2\sqrt{g_3 g_4}} \tag{9}$$

$$Z_o J_5 = \frac{\pi\Delta}{2\sqrt{g_4 g_5}} \tag{10}$$

The admittance inverter for the final coupling section is determined by

$$Z_o J_6 = \sqrt{\frac{\pi\Delta}{2\, g_5 g_6}} \tag{11}$$

Where Z_o is the characteristic impedance with value of 50 Ω , $(g_1, g_2, g_3, g_4, g_5 \text{ and } g_6)$ are the values for the elements of the fifth order Chebyshev with 0.5 dB ripple low pass filter prototype, J is the admittance inverter constant, and Δ is the fractional bandwidth with value of 0.1. The spacing between the coupled lines, length and width are calculated in ADS using the odd mode (Z_{0o}) and even mode (Z_{0e}) characteristic impedances of the coupled sections. For each coupling section, Z_{0o} and Z_{0e} are calculated by using the following equations:

$$Z_{0o} = Z_0[1 - JZ_0 + (JZ_0)^2] \tag{12}$$

$$Z_{0e} = Z_0[1 + JZ_0 + (JZ_0)^2] \tag{13}$$

Table 1 shows the calculated values of the odd and even impedances for the desired microstrip coupled line BPF [2].

3.6. Design and Simulation of a Low Noise Amplifier

The low noise amplifier (LNA) is used at the input of a receiver to amplify the low level signal with special care to minimize the effect of the unavoidable noise. The appropriate active device ATF-36163 is selected to design the LNA because it has a minimum noise figure and a maximum gain of 0.86 dB and 10.11 dB, respectively, at an operating frequency of 10 GHz. The ADS simulation is carried out to find the stability factor and determine the circles for the constant maximum available gain, and constant noise figure. Then the source reflection coefficient (⌜S) is selected at a point on the Smith chart that achieves an acceptable gain and noise figure according with the design goals. The load reflection coefficient is determined by the following equation.

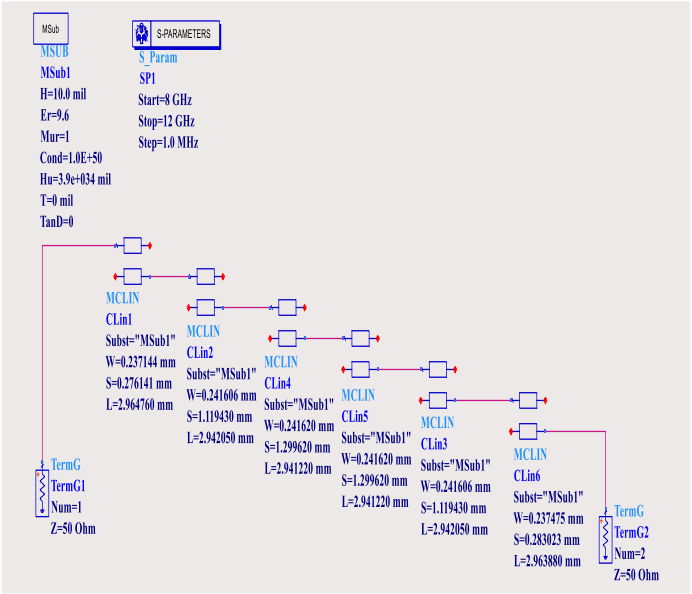


Figure 6: Simulation of a Microstrip Coupled Line Band-Pass Filter using ADS

Table 1: Calculation of Odd and Even Characteristic Impedances

Coupling section #1	$g_1 = 1.7508$	$Z_o J_1 = 0.13368$	$Z_{0o} = 44.21\ \Omega$	$Z_{0e} = 57.57\ \Omega$
Coupling section #2	$g_2 = 1.2296$	$Z_o J_2 = 0.02141$	$Z_{0o} = 48.95\ \Omega$	$Z_{0e} = 51.09\ \Omega$
Coupling section #3	$g_3 = 2.5408$	$Z_o J_3 = 0.01777$	$Z_{0o} = 49.13\ \Omega$	$Z_{0e} = 50.90\ \Omega$
Coupling section #4	$g_4 = 1.2296$	$Z_o J_4 = 0.01777$	$Z_{0o} = 49.13\ \Omega$	$Z_{0e} = 50.90\ \Omega$
Coupling section #5	$g_5 = 1.7058$	$Z_o J_5 = 0.02141$	$Z_{0o} = 48.95\ \Omega$	$Z_{0e} = 51.09\ \Omega$
Coupling section #6	$g_6 = 1$	$Z_o J_6 = 0.13571$	$Z_{0o} = 44.13\ \Omega$	$Z_{0e} = 57.70\ \Omega$

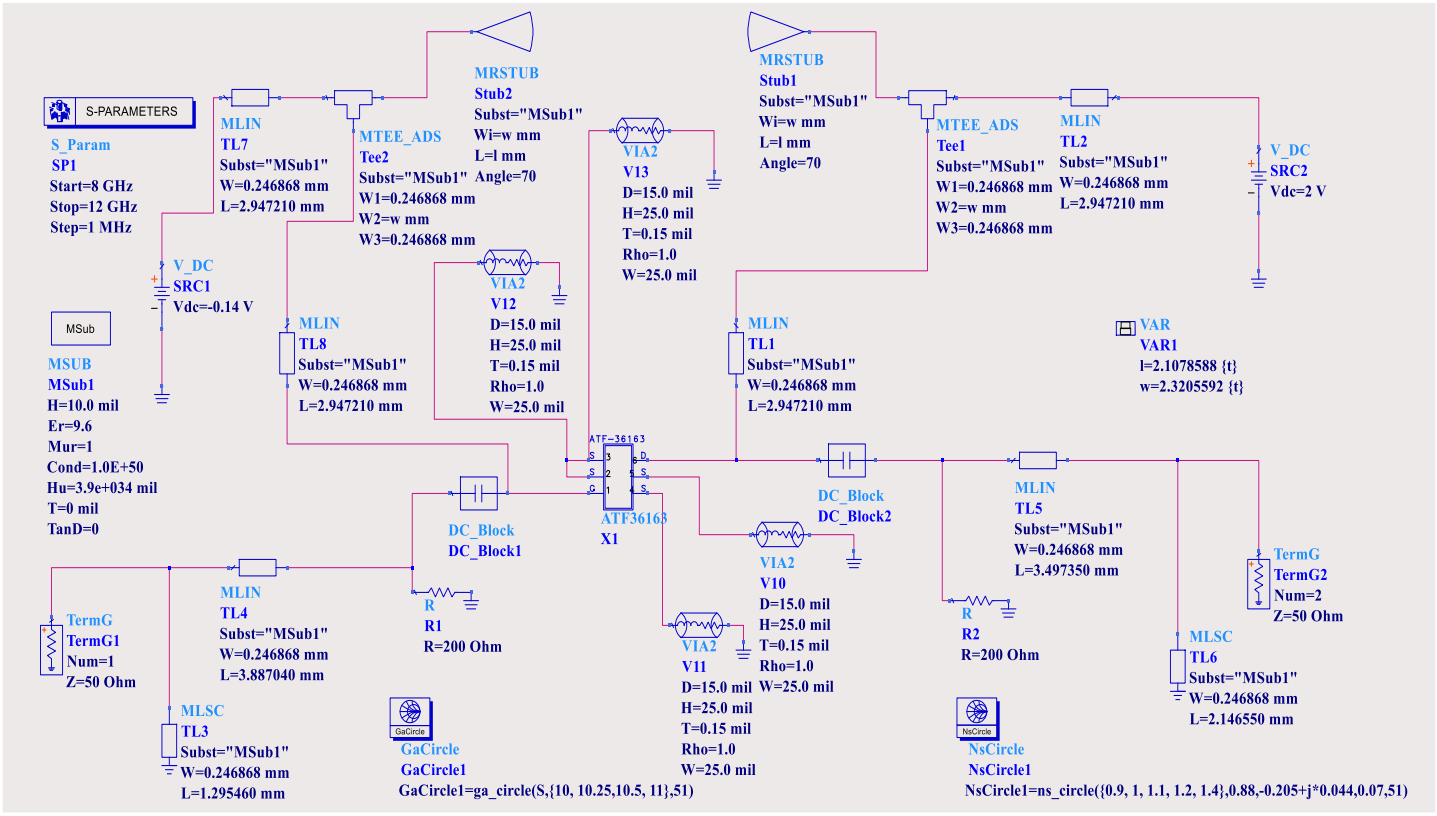


Figure 7: The Complete Circuit Design for the Proposed Low Noise Amplifier

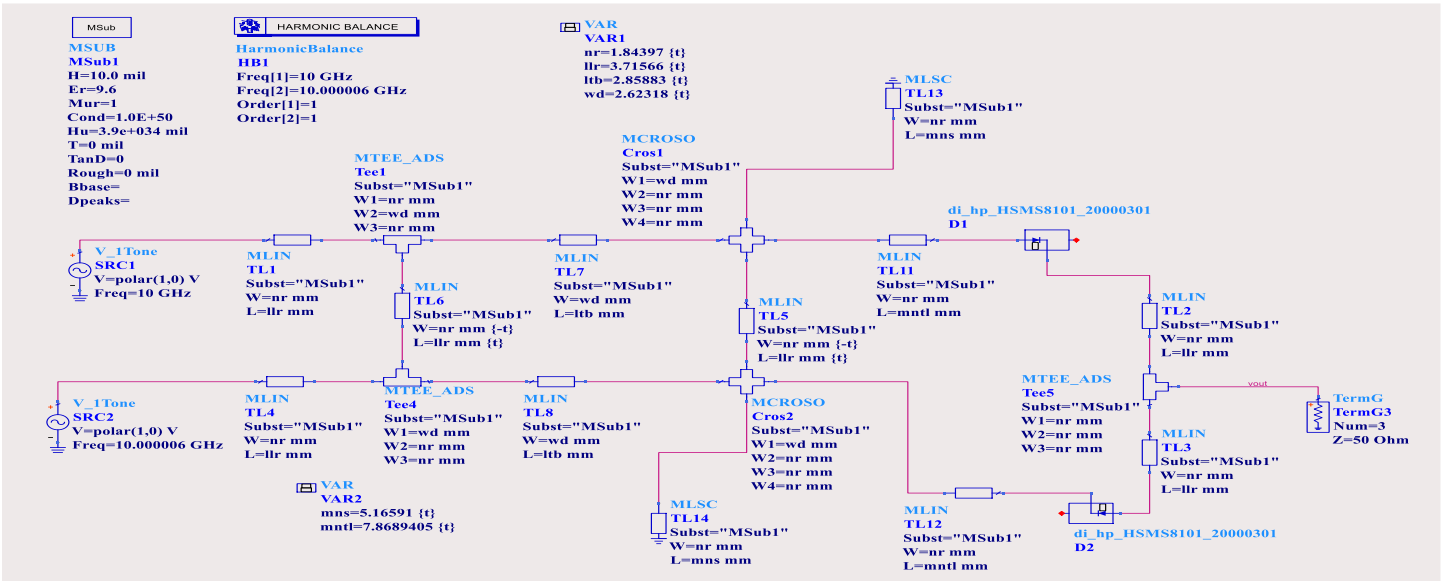


Figure 8: Simulation of a Microwave Single-Balanced Mixer using ADS

$$\Gamma_L = \Gamma_{out}^{*} = \left(S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1-S_{11}\Gamma_S}\right)^{*} \qquad (14)$$

The determined Γ_S and Γ_L are then converted to the source and load impedances, respectively. The input and output matching circuits are designed using microstrip transmission lines [3]. The radial stub is used to prevent the AC signals from entering the DC power supply. The series capacitors are used to block DC signals from entering the source and terminating tuning networks as shown in Figure 7.

3.7. Design and Simulation of a Microwave Single-Balanced Mixer

The proposed microwave single-balanced mixer consists of two microwave schottky diodes (HSMS-8101) which are connected via a quadrature hybrid. This mixer provides isolation between local oscillator (LO) and RF inputs. Either the LO or the RF signal is added destructively at the IF port of the mixer. The impedance matching is designed to convert the schottky diode's

impedance into 50 Ω. Figure 8 shows the schematic of the microwave single-balanced mixer [2].

4. Simulation Results

The harmonic balance (HB) simulator in ADS is used to perform the frequency domain analysis for the non-linear circuit of the proposed 10 GHz oscillator. The grounded oscillator port (Oscport) is used to predict the frequency spectrum and the time domain waveform of the oscillation circuit. Figure 9 and 10 show the frequency and time domains of the oscillation where the common gate transistor (ATF-13786) oscillates at 10 GHz.

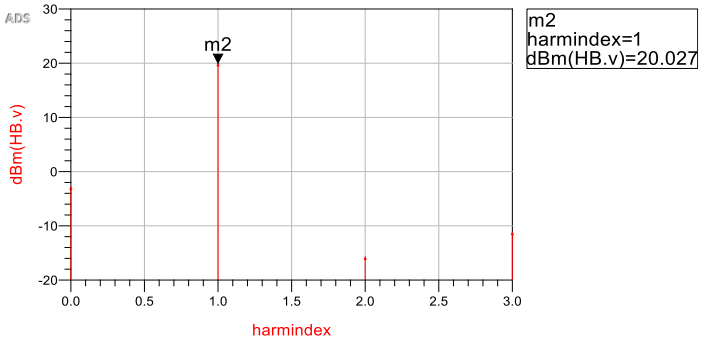


Figure 9: Frequency Domain Analysis for the Proposed 10 GHz Common Gate Oscillator

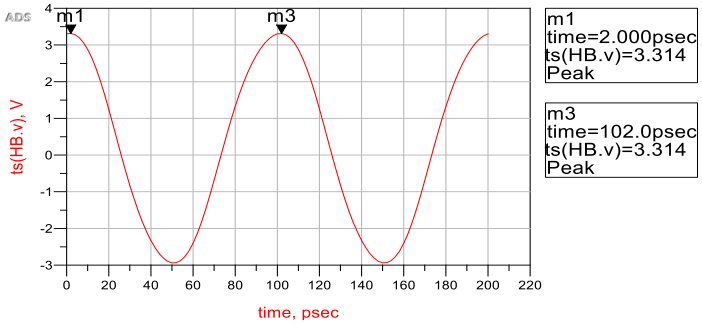


Figure 10: Time Domain Analysis for the Proposed 10 GHz Common Gate Oscillator

The S-parameters versus frequency for the Wilkinson power divider are simulated in ADS as plotted in Figure 11. At frequency of 10 GHz, the reflection coefficient at port 1 is small because the value for S_{11} is less than –40 dB; also the power equal-split between ports 2 and 3 is confirmed because the value of both S_{21} and S_{31} is about –3 dB.

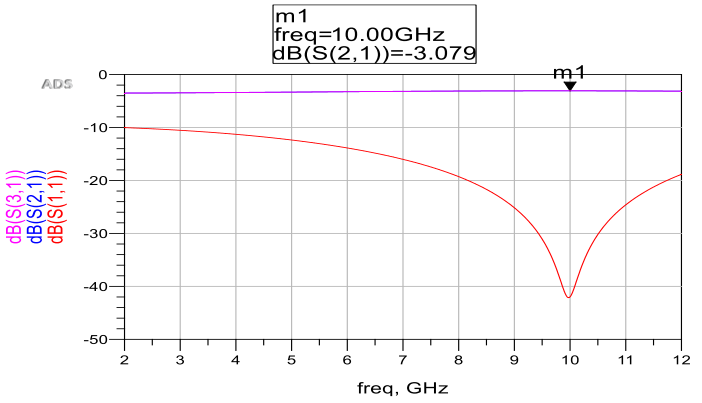


Figure 11: Simulated S-Parameters for Wilkinson Power Divider

The simulated S-parameters in ADS for the class-A power amplifier are shown in Figure 12. The complete circuit for the designed class-A power amplifier is well matched at 10 GHz because S_{11} is less than –30 dB. At a frequency of 10 GHz, the value of S_{21} indicates that the obtained gain at the output is 8.141 dB.

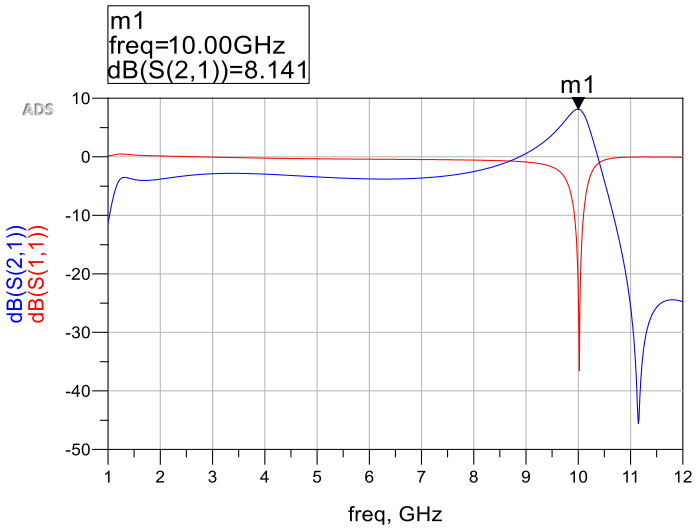


Figure 12: Simulated S-Parameters for the Proposed Class-A Power Amplifier

The CST MWS simulation results for the S-parameters and the three-dimensional radiation pattern for the designed Cantenna are shown in Figures 13 and 14, respectively. This Cantenna can achieve bandwidth of 15 % and gain of 7.755 dB at 10 GHz.

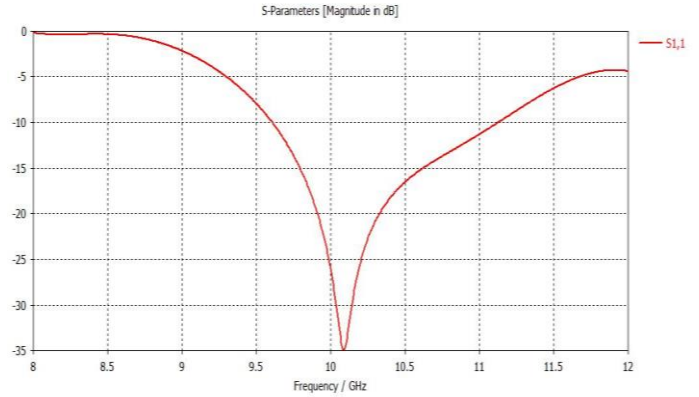


Figure 13: Simulated S-Parameters for the Proposed Cantenna

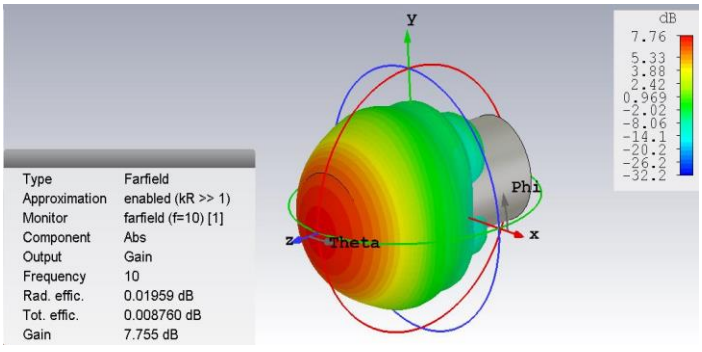


Figure 14: Three-Dimensional Radiation Pattern for the Designed Cantenna

The response of the microstrip coupled line band-pass filter is obtained using ADS simulation as shown in Figure 15. The center frequency is 9.97 GHz, while the band edges frequencies are 9.88 GHz and 10.06 GHz, respectively.

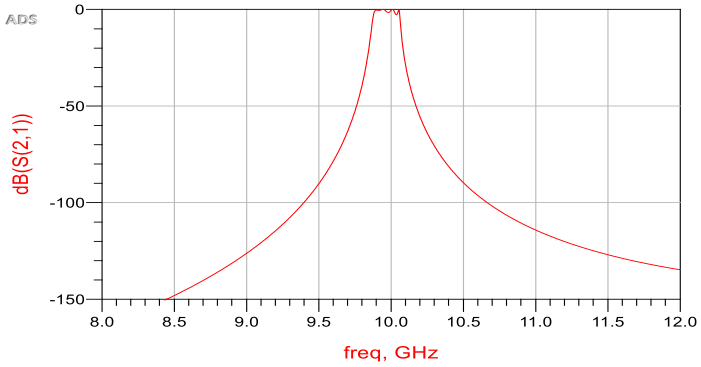


Figure 15: Simulated Response for the Microstrip Coupled Line Band-Pass Filter

S-parameters of the designed LNA are obtained by using ADS simulation as shown in Figure 16. At 10 GHz, simulations show a gain (S_{21}) of 10.488 dB, the circuit of the designed LNA is well matched because the values of S_{11} and S_{22} are –14.89 dB and –18.29 dB, respectively. Also the designed LNA circuit shows a good isolation because the value of S_{12} is below –16.44 dB.

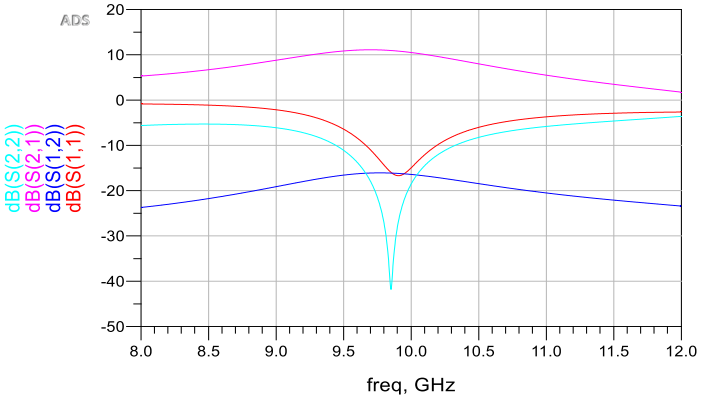


Figure 16: Simulated S-Parameters of the Designed LNA

The designed microwave single-balanced mixer is simulated in ADS. Figure 17 shows the spectrum at the output of the microwave single-balanced mixer; these results are obtained by using the harmonic balance (HB) simulator.

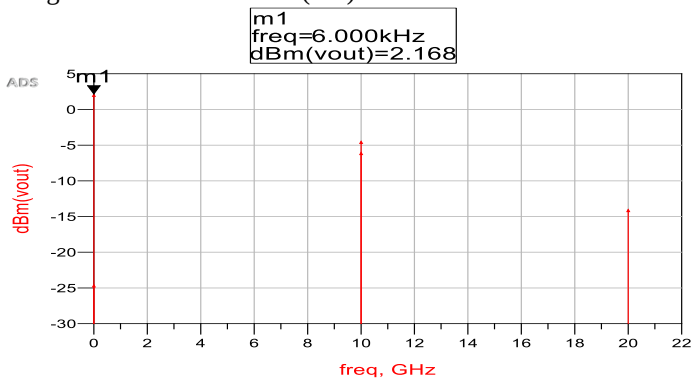


Figure 17: Simulated Spectrum at the Output of the Microwave Single-Balanced Mixer

The individual designed and simulated microwave components are connected together to form the complete RF Front-End transceiver for a continuous wave (CW) Doppler radar. This proposed Doppler radar RF front-end transceiver is simulated in ADS as shown in Figure 18. The amplified 10 GHz continuous wave hits an object. The received reflected wave will have a frequency higher than 10 GHz if the object moves toward the radar. On the other hand, the frequency of the received reflected wave will be lower than 10 GHz when the object moves far from the radar, this phenomenon is known as the Doppler effect. The reflected wave goes to the microstrip coupled line band-pass filter from which it is applied to the LNA, and then the output of the LNA is mixed with a portion of the transmitted wave.

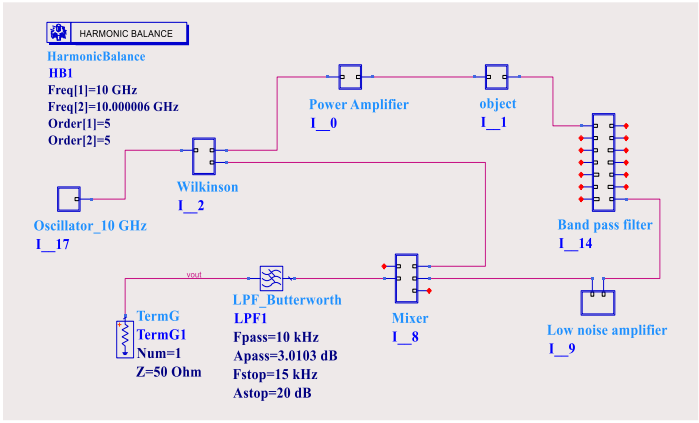


Figure 18: Complete Schematic of the RF Front-End Transceiver for a Doppler Radar

Figure 19 shows the ADS model of the object, this model simulates how the Doppler frequency is generated by the moving object. The frequency of transmitted continuous wave is shifted by a Doppler frequency ($\pm f_d$). So that the frequency shift in the transmitted continuous wave is simulated by multiplying this transmitted continuous wave with $e^{\pm j2\pi f_d t}$ or with $(\cos(\pm j2\pi f_d t) + j\sin(\pm j2\pi f_d t))$. The transmitted continuous wave is equally split by the power divider (PwrSplit2). One of the outputs of (PwrSplit2) goes through a phase-shifter (PhaseShiftSML) with phase shift of (90°) or (j) and then it is multiplied by $\sin(\pm j2\pi f_d t)$ using a multiplier (MULT2), the other output of (PwrSplit2) goes unchanged before it is multiplied by $\cos(\pm j2\pi f_d t)$ using (MULT1). The outputs of (MULT1) and (MULT2) are combined together using a power combiner (PwrSplit2).

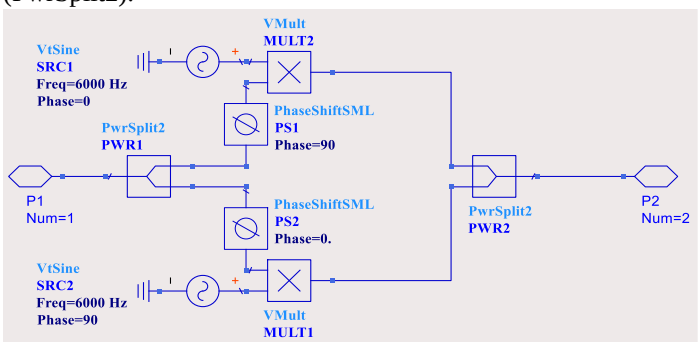


Figure 19: ADS Model of the Moving Object

The Doppler radar RF Front-End transceiver is simulated using ADS by selecting the Doppler frequency to be 6000 Hz as shown in Figure 19. The output of the microwave single-balanced mixer is filtered out using a low pass filter (LPF). At the output of the LPF, the frequency of the received continuous wave is confirmed to be equal the generated Doppler frequency as indicated in Figure 20.

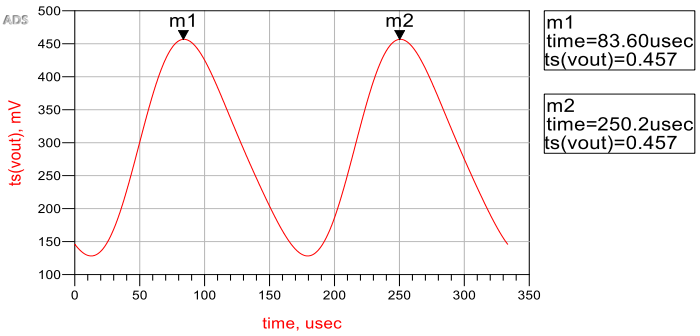


Figure 20: Time Domain of the Received Signal at the Output of the Low Pass Filter

5. Conclusions

Each microwave component of the continuous-wave radar RF front-end transceiver has been designed and simulated individually including 10 GHz oscillator, Wilkinson power divider, class-A power amplifier, Cantenna, microstrip coupled line band pass filter, low noise amplifier and microwave single balanced mixer. The microwave components have been connected together to form the complete schematic of RF front-end transceiver for a CW Doppler radar, then the proposed system has been simulated using ADS. This strategy in teaching the microwave engineering course uses the electromagnetics simulation tools such as ADS and CST and the theoretical knowledge in order to practice designing the microwave components.

Conflict of Interest

The author declares no conflict of interest regarding the publication of this paper.

References

[1] E. Aydin and N. E. Çagiltay, "Use of advanced technologies in a RF and Microwave Engineering course," *IEEE EDUCON 2010 Conference*, Madrid, pp. 811-817, 2010. doi: 10.1109/EDUCON.2010.5492495
[2] D. M. Pozar, Microwave engineering. Hoboken, NJ: Wiley, 2012.
[3] K.-W. Yeom, Microwave circuit design: a practical approach using ADS. New York: Prentice Hall, 2015.
[4] Teng, Kevin Chi-Ming, "The Design and Evaluation of a 5.8 GHz Laptop-Based Radar System" Master Thesis, Purdue University, 2013.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 258-267 (2019)

www.astesj.com

ASTESJ

ISSN: 2415-6698

CISELexia: Computer-Based Method for Improving Self-Awareness in Children with Dyslexia

Aulia Rahmawati*,[1], Emil Robert Kaburuan[2], Anditya Arifianto[1], Nahda Kurnia Juniati[3]

[1]*School of Computing, Telkom University, Bandung, Indonesia 40257*

[2]*Information System Management Department, BINUS Graduate Program - Master of Information System Management, Bina Nusantara University, Jakarta, Indonesia 11480*

[3]*Psychology, Informatics and Business University, Indonesia 40286*

ARTICLE INFO

Article history:
Received: 30 July, 2019
Accepted: 13 September, 2019
Online: 08 October, 2019

Keywords:
Dyslexia
Emotional-intelligence
Self-awareness
Gamification

ABSTRACT

Dyslexia has a lack of ability which causes other disturbances to social-emotional development. Based on an interview with the Chair of the Indonesian Dyslexia Association, increasing self-awareness is one form of dyslexia therapy to reduce that problem. Therefore there is a need for increasing self-awareness to solve that problem, whereas such technology the closest has not been fit with dyslexic children need. Then this study tries to focus on designed and developed CISELexia (Computer-Based Method for Improving Self-Awareness in Children with Dyslexia) technology using a gamification approach. It aims to help dyslexic children improve their competencies by increasing their self-awareness. Based on the implementation results CISELexia can increase the child's self-awareness state according to the rubric questionnaire given for about 10%.

1. Introduction

Dyslexia is a type of language disorder which further affects reading, writing, spelling, and organizational skills. Early diagnosis of dyslexia involves disruption to academic abilities, such as problems with reading, text comprehension, spelling, writing, counting, and mathematical reasoning. With this lack of ability, it can cause other disturbances to social-emotional development. Based on an interview with the Chair of the Indonesian Dyslexia Association, it was stated that dyslexic children have a tendency to make mistakes in understanding their emotions (Dewi K., 2017). Increasing self-awareness should solve that problem. Unfortunately, the closest technology available for social treatment has not been fit for the needs of dyslexic children.

Dyslexia cannot be cured, however, the symptom can be reduced. One of them is through therapy. And along with the development of technology, some technologies can help the process of this therapy such as the use of video, games, and movies [1–8]. The closest technologies that can overcome that problem are SIDES (Shared Interfaces to Develop Effective Social Skill), AVP (Authorable Virtual Peer), computer-presented social stories and video models. But, those technologies are designed for autism child, which those has different characteristics with a dyslexic child. The lack of technology that can support therapy on social-emotional development, make an opportunity to start this research.

As previously mentioned, there is no cure for dyslexia, then one of the solutions that have been offered by researchers was providing therapy to treat people with dyslexia. This therapy is usually done by a professional in a clinic-based session. But along with the development of technology, various ways can help the process of this therapy such as the use of video, games, and movies [1–8]. One of the technologies that support the therapy of academic skills in dyslexia is Lexipal. Lexipal is a technology that uses multisensory approach combined with gamification approach to build "learn-to-read" therapy for children with dyslexia. The advantage of this technology is that Lexipal is made to meet the needs of Indonesian users. So, it can be accepted by users in Indonesia. The development of this technology has been adapted to some aspects of academic skills such as reading, writing, and counting [8].

Several examples of proven technologies which can be used in social skills improvement therapy are SIDES, AVP, computer-presented social stories and video models [9–13].

* Aulia Rahmawati., Email: auliasuwito@gmail.com

Some of the technologies mentioned earlier, such as Lexipal and SIDES use game-based concepts as its core. This proves that game technology has the advantage of enhancing the treatment [2, 3, 6, 8, 10]. The same thing also revealed by some researchers that there is game technology that is used to help dyslexic children such as iLearnRW, Dyslexia Baca and Dyseggxia [2, 3, 6]. Those products are the examples of using gamification approach in developing dyslexia applications. Gamification is using game elements in non-game context purposes. Like Lexipal, Dyseggxia, Dyslexia Baca and iLearnRW which use game elements to build their application for learning purposes.

Based on the existing problems where the lack of technology-supported therapy for social-emotional development, this research tries to design and develop a technology to help dyslexic children to increase their self-awareness. The technology that will be designed and developed is CISELexia (Computer-Based Method for Improving Self-Awareness in Children with Dyslexia) which uses gamification approach and child-centered designed method to help develop the system that is by the stages of social-emotional development and dyslexic children needs.

2. Related Works

2.1. Theory of Dyslexia

Some references have mentioned the characteristics of dyslexia, including the American Psychiatric Association, International Dyslexia Association, Cuschieri, Solek, Marshall and Pennington. The American Psychiatric Association and the International Dyslexia Association describe the characteristics of specific learning difficulties, types of specific learning difficulties and diagnostic criteria for specific learning difficulties. On the other hand, Abigail Marshall describes dyslexia difficulties, how to teach dyslexic children and how to overcome the educational, personal and social problems. The difficulties that present in dyslexia sufferers are:

1. Inaccurate or slow word reading
2. Difficulty in understanding the meaning of the text
3. Difficulty in spelling
4. Difficulty in writing the expression
5. Difficulty in mastering numerical sense, number fact or calculation
6. Difficulty in mathematical reasoning

A person with dyslexia has normal or above-average intelligence. Dyslexia affects not only academically but also daily life aspects, one of which is emotional intelligence. Studies from [14] resulted that behavioral disorders and social problems occur more in dyslexic children than in non-dyslexic children. Several studies have shown that children with learning disabilities show higher aggression and misbehavior [15]. This shows that they have low emotional intelligence. [16] describes emotional intelligence: “Emotional intelligence is the ability to perceive emotions, to access and generate emotions to assist thought, to understand emotions and emotional knowledge, and to reflectively regulate emotions to promote emotional and intellectual growth.”

This emotional intelligence is one of success indicators in social-emotional development. Emotional intelligence is one of the factors that influences social-emotional development. A study says that dyslexic children have social-emotional problem because of the relatively low level of emotional intelligence [17].

Daniel Goleman describes the five components of emotional intelligence:

1. Self-awareness
2. Self-regulation
3. Internal motivation
4. Empathy
5. Social competence

Self-awareness is the ability to recognize and understand moods, emotions and personal impulses, and their influence on others. Social emotions depend on the ability of self-awareness to monitor the emotional state and identify to name that emotion.

Self-awareness includes the ability [18]:

1. Emotional self-awareness
   Recognizing one’s emotion and their effects
2. Accurate self-assessment
   Knowing one’s strengths and limits
3. Self-Confidence
   A strong sense of one’s self-worth and capabilities

2.2. Dyslexia Treatment

According to Goleman and Indigrow Child Development Center Indonesia (Goleman, 1996a), there are two treatments performed on dyslexia sufferer: academic and social treatment or known as a social-emotion treatment. Treatment is performed in early childhood or about five years old (preschool) for maximum results [20]. Previously treatment in dyslexic children done by professionals in clinic-based sessions. But now the technology has begun to be developed to support treatment in children with dyslexia [1, 2, 11–13, 3–10].

2.2.1. Academic Treatment

Academic treatment focuses on developing academic skills such as reading, writing, and counting. One of the technologies that support the therapy of academic skills in dyslexia is Lexipal. Lexipal is a technology that uses multisensory approach combined with gamification approach to build "learn-to-read" therapy in children with dyslexia. The development of this technology has been adapted to some aspects of academic skills such as reading, writing, and counting [8]. Other technologies that support academic treatment are Dyslexia Baca and Dyseggxia [3, 6].

2.2.2. Social Treatment

Social-emotion treatment focuses on other skills that support social skills improvement. According to Goleman (Goleman, 1996a), social-emotion has several competencies that begin with self-awareness, self-regulation, motivation, empathy and social competence. Social competence is the highest stage in social skill, that is all forms of skill involving social behavior.

Since this study focuses on self-awareness, there are three competencies in self-awareness: emotional self-awareness, accurate self-assessment, and self-confidence [18]. From that competency, this research, indicators divided into five representing self-awareness ability. The indicators are:

1. *Know yourself*
   This indicator represents how much we know about self-identity such as name, date of birth, age and class.
2. *Know your family*
   This indicator represents how much we know about our families like the names of parents, the number of brothers, the number of sisters, and their names.
3. *Know your strengths*
   This indicator represents how much we know about our strengths. This research uses soft and academic skills for representing it.
4. *Know your weakness*
   This indicator represents how much we know about our weaknesses. This research uses soft and academic skills for representing it.
5. *Know your emotion*
   This indicator represents how much we know and understand each other's emotions and ourselves. Emotions that need to be mastered according to Daniel Goleman are categorized into eight emotions, they are anger, sadness, fear, happy, love, shock, annoyed and shame. Treatment on this indicator is done by studying the emotions of the other through pictures or videos of social situations.

2.2.3. Existing Social Treatment

According to observation in Indigrow Child Development Center, the existing social treatment is done by viewing social stories using video, then the child is asked to respond to what happened in the video. Psychologists also provide some questions related to the social situation based on the video. According to psychologists, within three months the children have been increasing self-awareness skills.

While there is no special assessment conducted to find out the self-awareness of children. Existing assessment is done only based on observations from the interactions between psychologists and children.

2.3. Previous Works

Some technologies have proven to improve social skills such as SIDES and AVP [9–13]. SIDES (Shared Interfaces to Develop Effective Social Skill) is a technology designed to assist Asperger's syndrome in training group work skills. This technology has form as tabletop computer games that contain case studies to solve in groups with the aim support effective group work created. It involves highly visual and uses cards, pattern building, and puzzle-solving as game mechanics. They say that cooperative computer games are effective to teach group work skills [10]. There is also an Authorable Virtual Peer (AVP), a technology that helps children learn to interact with each other in order to improve social skills through learning to interact with virtual peers. This technology is designed using a shared reality concept, dialogue, and simulation as game mechanics. It is addressed for autism spectrum disorders that have peer social interaction problems. AVP provides a lesson on how to interact with peers [13]. Another technology is by using computer-presented social stories and video models can have a positive effect on teaching the social skills of children with autism syndrome. They use social stories and video models to make an intervention on children with autism [12].

2.4. Theory of Gamification

The gamification approach is the concept of using game elements in a non-gaming context to improve user experience, learning, problem-solving, motivation and user engagement. The examples of non-gaming contexts are marketing, politics, education, health, and fitness. Using gamification can be a solution to achieve excellent training goals. This is proven by the use of gamification in game-based learning in some previous research [21–25]. To achieve those goals, it is necessary to know the game elements that are capable of supporting three dimensions of learning: cognitive, emotional and social [26–28].

In the cognitive area, gamification can use challenges tailored to player skills. In the emotional area, it involves the emotion of the user, how the gamification can trigger user emotions such as joy, curiosity, and frustration. And in the social area, gamification can provide a social experience like experience in new identities and roles, so the user can make a decision according to point of view [27].

One example of the use of the gamification approach in dyslexia treatment is Lexipal. Lexipal uses some game elements such as story, level, point, reward, and achievement. The use of gamification in Lexipal is for increasing motivation in learning [8]. The other examples are iLearnRW, Dyslexia Baca and Dyseggxia [2, 3, 6].

3. System Design

3.1. Application Design

The design of CISELexia application is tailored to the characteristics of dyslexic children. Some of the characters of dyslexic children which are considered to make application design including:

1. Difficulty in spelling letters.
2. Difficulty in making written work such as essays.
3. Poor memory.
4. Difficulty in understanding sentences that are read or heard.
5. When listening to something, the attention span is short.
6. Difficulty in remembering names.
7. Difficulty in visual discrimination.

Based on research on the characteristics of dyslexic children, the results of the CISELexia application interface plan are as described in the following points.

Main menu Scene

Figure 1: Main Menu Scene

In the main menu scene, the menu buttons are arranged vertically. The play button is made larger than the other buttons to attract the attention of children to try it first.

Home Scene

Figure 2: Home Scene

Before the children can start trying out the existing tasks, the children are introduced first to each character that appears on the task so that they can easily follow the storyline. Dyslexic children sometimes have difficulty in visual discrimination, so the buttons to start the task are made in various shapes and colors with the aim of introducing various shapes to the children. Compared to giving the task name to the text button, the text button contains a sentence that makes the children curious and want to press the button. The font used in CISELexia also uses a special font adapted to dyslexic children, namely OpenDyslexic font. The placement of the home button is on the upper left side and the settings button on the upper right side. The placement of these two buttons is consistent in all scenes, so this makes it easy for the children.

Know Yourself and Know Your Family Scene

Figure 3: Example of Know Yourself/Family Scene

When the children press a button that leads to the *know yourself* or *know your family* scene, a video will appear describing the condition of the social situation with a narrative that can explore the children's identities and family through questions that appear in the middle of the video. Random choices are used to avoid children memorizing the pattern of question and answer.

Know Your Strength and Know Your Weakness Scene

Figure 4: Example of Know Your Strength/Weakness Scene

Know your strength and *know your weakness* scene, the system provides a list of images about academic skills, sports, art, and daily activities. Furthermore, the child is asked to choose images according to their strength and their weaknesses.

Know Your Emotion Scene

Figure 5: Example of Know Your Emotion Scene (First Figure)

The design of this scene is divided into two. The first scene is for emotion identification through images, see figure 5 and the second is emotion identification through video, see figure 6. The video scene is almost similar to the *know yourself* and *know your family* scene. That is when the children press the button that leads to this scene, a short animated video will appear and at the end of the video, there are some questions about emotions shown by the characters. Animated videos are made in short duration to overcome the short attention span of the children. Random choice is used to avoid children memorizing the pattern of question and answer.

Figure 6: Example of Know Your Emotion Scene (Second Figure)

3.2. Video Content Generation

This video content is made to *know yourself*, *know your family* and *know your emotion* indicators. For *now, yourself* and *know your family* indicators, the use of video is intended to create a social situation that can help children to recognize themselves and their families. While for the *know your emotion* indicator, the video is intended to help children understand emotions in everyday social situations. Narratives for video content is made so that it does not cause misperception in understanding emotions. One example of the narrative is as follows.

Example 1:

"That morning, I, my sister and my brother were reading a book in the living room. We read different books. 'Sis, can I borrow the book?' said my brother, pointing at the book held by my sister. 'No you can't, I am still reading it,' replied my sister.

Because it is not lent, my brother immediately seizes that book and then tore it. My sister looked at my brother cynically."

Example 2:

"It was already 6:00 in the morning. My mother screamed to wake me up. 'Andy wakes up, it's already 6 o'clock, you will be late for school.' But because I was still sleepy, I didn't respond to my mother's call and continued to sleep. The time showed 06:30, my mother screamed louder and began to go to my room.

'Son wake up quickly!!', My mother's voice was louder and she tried to shake my body. Then I woke up. I woke up lazily. 'Wake up kid, it's too late! Your pickup bus has departed! You must be late for school.'"

The psychologist said that two examples above are narratives that can be used on video content for your know emotion indicator. But for example 2 is an example of an unfavorable narrative for emotional video content because it contains ambiguous emotions, that is, the mother's emotions being angry because the brother/sister who is late getting up and the mother's emotions being sad because the shuttle bus has departed. The narratives must pass an evaluation process from a psychologist to get video content that is appropriate for the target audiences.

The expected video results are far from multiple perceptions. So that in this research, we use 3D technology that will make it easy to set the camera to get the right viewpoint, so multiple perceptions video can be avoided.

3.3. Character Design

In making the characters for video content, this study utilizes 3D technology. The character formed is also adjusted to the characteristics of children with dyslexia. That is by creating unique characters, easily recognizable and unambiguous. Such as the use of hijab or skirt to depict female characters, and used shorts with short haircuts to depict male characters.

Figure 7: Example of female characters (Teacher)

3.4. Task Generation

Based on the background of this research, CISELexia was proposed as the technology solution to improve self-awareness. Thus CISELexia is designed based on indicators in self-awareness. This indicator is used as a parameter for measuring self-awareness increase. The indicator can be seen in figure 8.

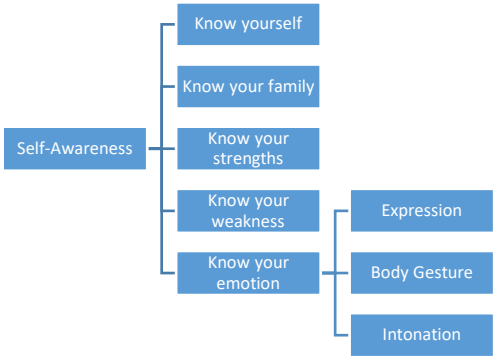


Figure 8: Indicators of Self-Awareness

Based on existing indicators, the tasks are made on the CISELexia system that represents each indicator. The tasks which are developed in the CISELexia application are adapted to the indicators that will be achieved based on self-awareness theories which have been studied previously. This task is divided into 7 types and represents 5 indicators, can be seen in Table 1 for more details.

Table 1: Mapping Table for Tasks with Indicators

	Know yourself	*Know your family*	*Know your strength*	*Know your weakness*	*Know your emotion*
Task 1	√				
Task 2		√			
Task 3			√		
Task 4				√	
Task 5					√
Task 6					√
Task 7					√

- *Know yourself*

  For *know yourself* indicator, it will be accommodated by task 1, where there are dialogue and some exercises in the form of questions about identity such as name, age, and class.

- *Know your family*

  This indicator will be accommodated by task 2, by displaying a dialogue in the form of a social situation which involves exercises in the form of questions about our family, such as the names of parents, the number of brothers and the number of sisters.
- *Know your strengths*

  This indicator will be accommodated by task 3, the system will provide various types of strength and the user will choose the strength that suits him. Where types of strength are used based on soft and academic skills.
- *Know your weakness*

  This indicator will be accommodated by task 4, the system will provide various types of weakness and the user will choose the weakness that suit him. Where types of strength are used based on soft and academic skills.
- *Know your emotion*

  These indicators will be divided into three tasks that are accommodated by task 5, 6 and 7. In task 5 the system provides various images of the social situation for analyzed by the user and what should we do. In tasks 6 and 7 system provided a social story in the form of video and then the user will be asked to answer the question that displayed.

Each training task has its own game mechanics with the following details:

1. Task 1-2

   Game mechanics in these tasks is a social story with the visual novel game model, which is a game with a fiction interaction game featuring a story and comes with a conversation as a narrative (dialogue) of each character. Among the dialogues that occur between actors, will be available questions that must be answered by the user as an assessment process on the indicator.
2. Task 3-4

   Game mechanics in these tasks is that the application provides some choices that represent strength and weakness, then the user must choose some choices that represent themselves.
3. Task 5-7

   Game mechanics in these tasks is almost similar to tasks 1-2 which is the visual novel model. The difference between them is the question of the assessment process. In this task, the user is asked to understand and analyze the social situation that occurred. So that the user will understand the emotions of the characters that exist in the situation, hope the user can answer the question that posed.

The whole question in the task is multiple choice. The system provides answer options, so the user only needs to choose the right answer, the user does not need to type their answer.

3.5. The Use of Gamification Approach

In order for CISELexia could be used to intervene in increasing self-awareness capabilities, the gamification approach is inserted in the application. The gamification approach is the concept of using elements of game design in a non-gaming context to improve user experience, learning, problem-solving, motivation and user engagement. Gamification on the CISELexia application can be seen from the scores, stars, and feedback. The use of score and star systems can be seen in figure 9. With the score, the system can trigger children to always increase score in each application usage. While the use of feedback is in the voice form at the end of the answering process. The feedback system can also be used to trigger children to find out their mistakes and try to fix them.

Figure 9: Example of Score and Star Systems

3.6. Scoring System

The assessment used in the CISELexia system is divided into two types. The first, for *know yourself*, *know your family* and *know your emotion* indicators using the formula on equation 1.

$$score = \frac{number\ of\ correct\ answers}{number\ of\ questions} \quad (1)$$

The second, *know your strength* and *know your weakness* indicators using the formula on equation 2.

$$score = \begin{cases} 0, if\ B' \cap A \\ 1, if\ A' \cap B \\ 2, if\ A \cap B \end{cases} \quad (2)$$

With is a strength/weakness list based on parents or psychologists assessment and is a strength/weakness list based on the child's answer.

4. Implementation and Result

4.1. Implementations

This implementation was carried out by dyslexic children along with their psychologists. Children will be asked to use the system as a form of exercise in recognizing self-awareness abilities, while psychologists are asked to review children's self-awareness abilities before using the system and after using the system during certain intervals.

4.1.1. Implementation Objective

In accordance with the objectives of the research mentioned earlier that this study aims to investigate the effect of using technology in order to support dyslexic children in improving self-awareness. Therefore, in this implementation, the historical record of the children was made when they use the system.

4.1.2. Implementation Strategy

The implementation strategy carried out in this study was doing test in dyslexic children with one-month intervals, the use of the system was done once a week. In addition, in order to be able to answer the hypothesis in this study, a questionnaire was taken using the pre-test and post-test rubric by the psychologist to

see the child's self-awareness abilities before and after using the application.

The initial implementation in this study was conducted by collecting personal data of children identified with dyslexia with an age range of 7-13 years. The initial implementation was carried out on five dyslexic children.

This child's personal data will be used as the child's answer validation when using the CISELexia system. In addition, in this study, there are also three other types of data, namely pre-test data, answer data, and post-test data.

Pre-test data is an individual respondent's personal review data on self-awareness ability before using the CISELexia system. This data is obtained from a children's self-awareness rubric questionnaire. The answer data is the respondent's historical data taken from a collection of individual answers while using the CISELexia system. The post-test data is the respondent's individual personal review data on self-awareness aspects after using the CISELexia system. This data is also obtained from children's self-awareness rubric questionnaire.

In order to achieve the objectives of the research, the strategy carried out in this implementation is to test the CISELexia system for chosen respondents during certain intervals. In this implementation, it was carried out in 1-month intervals with 1 use each week. Then the results of changes from respondents who use the system compared with the average time needed by respondents to be able to achieve changes in the aspect of self-awareness with a model without a system.

4.2. Implementation result

Based on the implementation results obtained data of a child's self-awareness conditions from a psychologist before application usage (pre-test), during the application usage (1st-obs, 2nd-obs, 3rd-obs, and 4th-obs) and data of child's self-awareness conditions from a psychologist after application usage (post-test). From this data, the data were categorized into 5 indicators, *know yourself*, *know your family*, *know your strength*, *know your weakness* and *know your emotion*. The range of values used in this data is 0-1. With a value close to 1 means the child is increasingly mastering the indicator.

The implementation shows the average child has improved their self-awareness state based on the rubric questionnaire in this research. To see the whole result of the implementation, it can be seen in table 2.

Table 2: Average score of self-awareness indicators in whole implementation result

Indicators	Pre-Test (psychologist)	1st Obs	2nd Obs	3rd Obs	4th Obs	Post-Test (psychologist)
Know yourself	0.973	1	1	1	1	1
Know your family	0.9	0.95	0.95	0.95	1	0.934
Know your strength	0.4	0.546	0.619	0.673	0.728	0.6
Know your weakness	0.4	0.291	0.364	0.4	0.51	0.6
Know your emotion	0.668	0.756	0.834	0.934	0.967	0.734
Average	**0.668**	0.709	0.753	0.791	0.841	**0.774**

The results of the study show that the average child has mastered *know yourself* indicator. It was seen that the child was able to complete the task from the application. The data from the psychologist also stated the same thing that the children increase their score from 0.973 to 1.

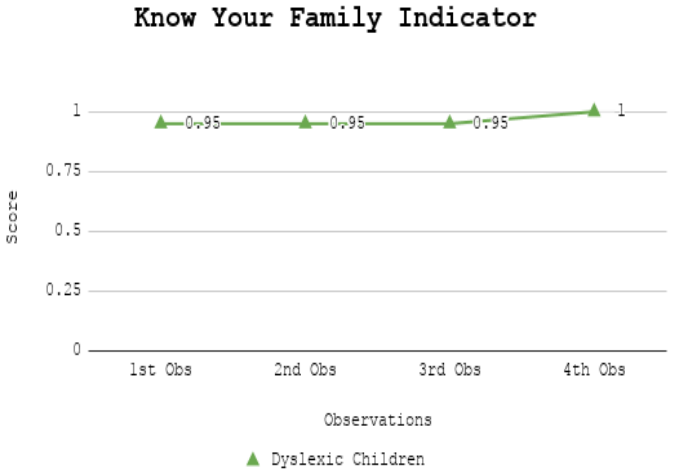


Figure 10: Average Results of Know Your Family Indicators Assessment

The average child has mastered this indicator. Seen at the end of the observation the child is able to complete all the tasks and this also shows an increase compared to previous observations. The same thing was also shown from psychologist data, which has increased from 0.9 to 0.93. But even though, there are some children who are still having difficulty distinguishing siblings or not. That's why at the beginning of observation the values achieved are not perfect and the increased score from psychologists is low.

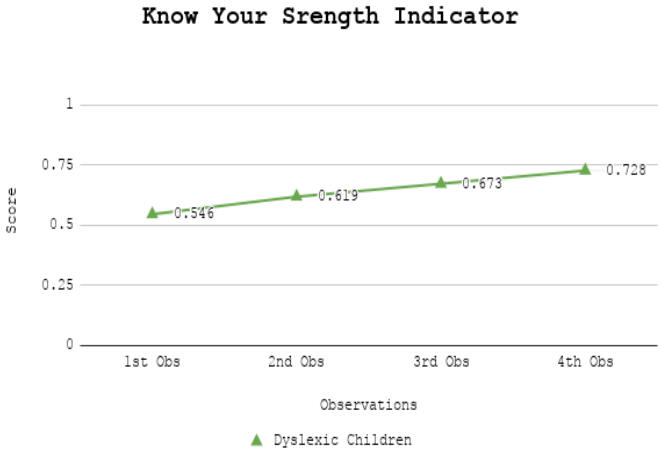


Figure 11: Average Results of Know Your Strength Indicators Assessment

Overall the *know your strength* indicator chart shows an increase during application usage. The results from the psychologist also showed an increase, based on table 2 that is from 0.4 to 0.6. But in this case, the psychologist states that the child's ability to express his strength is quite difficult to master, thus causing a not too significant increase and some even constant, one of them can be seen in figure 12.

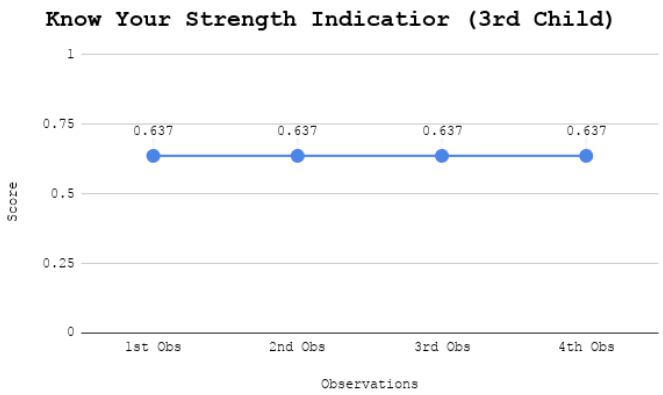


Figure 12: Assessment Result of Know Your Strength Indicator for 3rd Child

Above is an example of a child who does not give an increase in the *know your strength* indicator at each observation session. They are still inconsistent, and sometimes they give different answers in each session.

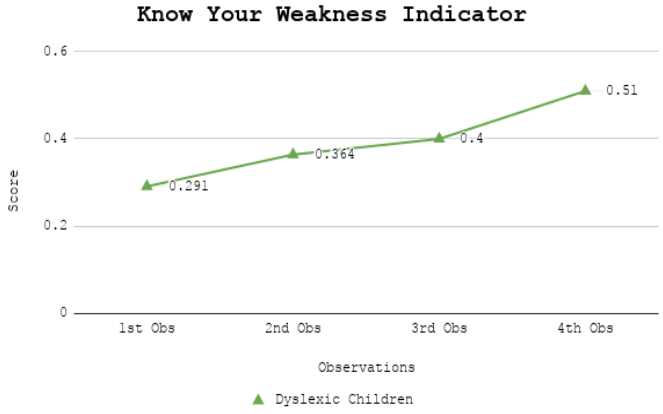


Figure 13: Average Results of Know Your Weakness Indicators Assessment

For the indicator of *know your weakness*, it is an indicator that is quite difficult to be mastered by children. Therefore the average increase value is small (only ±0.073) and not significant. It could be because of the inconsistency of children in answering their weaknesses in each observation session. There are constant values, some even go down, and one of them can be seen in figure 14.

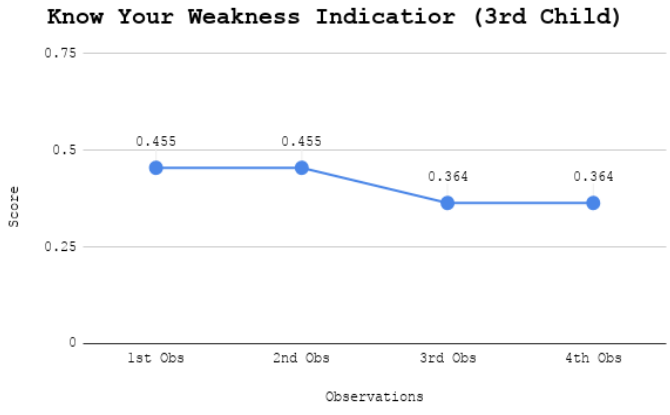


Figure 14: Assessment Result of Know Your Weakness Indicator for 3rd Child

Above is an example of a child who does not give an increase in the *know your weakness* indicator at each observation session. This case like on *know your strength* indicator, that the child still inconsistent, and sometimes he gives different answers in each session.

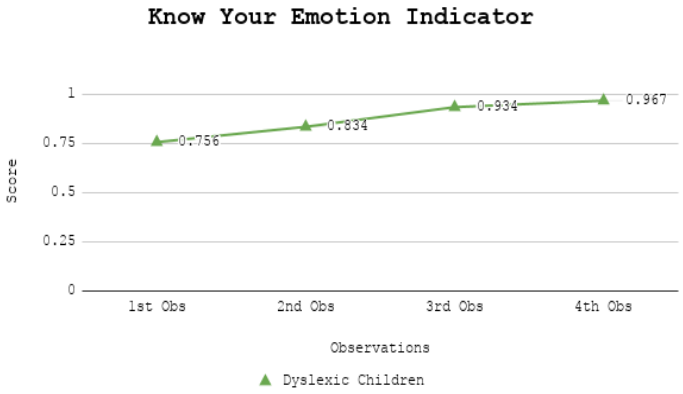


Figure 15: Average Results of Know Your Emotion Indicators Assessment

The implementation results show that the *know your emotion* indicator also increases for about 0.211 from the first observation 0.756 and the forth observation 0.967 as shown in figure 15. But in this case, the psychologist data has a different state than the application, from the psychologist the increased score for about 0.067 that is from 0.668 to 0.734. It means, the application almost states children have mastered this indicator, while psychologists say there are some points that are still not mastered by children. This can be seen in figure 16 where there are children with the same pre-test and post-test values on this indicator.

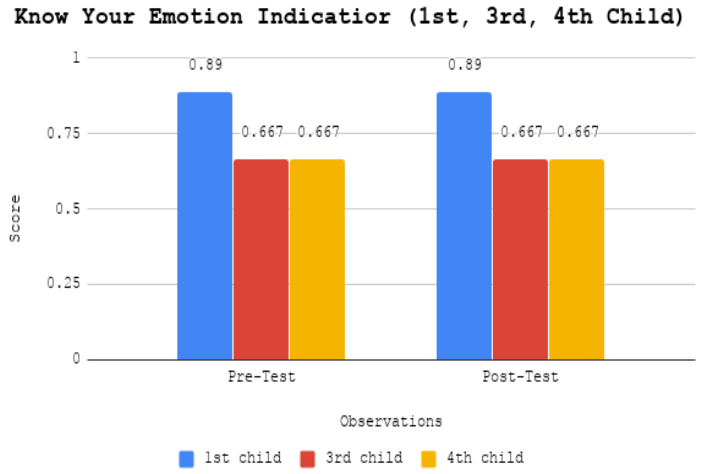


Figure 16: Assessment Result of Know Your Emotion Indicator for 1st, 3rd, and 4th Child

From all of the data, it can be concluded that the average score of rubric questionnaires given for about 10% that is from 0.668 to 0.774. A conclusion that can be derived from the data is that the data is a comparison to show an increase in self-awareness value of dyslexic children based on rubric assessment. Percent increase is not a benchmark to know the good or bad of the increase obtained. This is because this research is the first research and there is no supporting research that can be used as a reference.

4.3. Discussion

Implementations conducted to identify how technology gives the influence to help dyslexic children improve their self-awareness abilities. Based on the analysis of the implementation results, there are several things that can be discussed:

1. In making applications for children with dyslexia or similar children, things that need to be considered and concluded from this research are the creation of learning content, the creation of character, making tasks for learning content and additional details that can increase children's interest in learning.
2. In the testing application that has been made, for applications with specific target users, it is recommended to go through testing and evaluation by experts. This is done to get results that are more targeted.
3. For some indicators such as *know yourself* and *know your family*, most children have mastered this indicator, even though there are children who have not mastered it, at the end of the observation session shows an improvement. This might happen because the system provides feedback when the child gives the wrong answer, so the child can be triggered to find out things that they have not been mastered. This also been proven by [8] that feedback gives positive responses thereby minimizing their mistakes.
4. For the *know your strength* and *know your weakness* indicator, it is a difficult indicator to be mastered by children. This can be seen from the inconsistency of children in answering and the results of the pre-test/post-test that did not increase significantly and some even constant.
5. For the *know your emotion* indicator, from the results of system usage the child has a tendency to increase self-awareness ability, but in the assessment of pre-test/post-test by psychologists this is not appropriate. Most children do not give an increase in these indicators. This might happen because they helped by the system by giving feedback so they might remember to do not make the same mistakes and be motivated to improve their answers. Besides that, the use of points and rewards can also provide encouragement and motivation during the treatment process. This is also proved by [8]. This was also done during treatment with a psychologist, they always gave feedback on their answers. But in reality children still have not mastered this ability during the research period. In traditional treatments, children usually need approximately three months to be able to make an improvement in this indicator.
6. Based on the implementation results overall CISELexia can increase the child's self-awareness state according to the average score of rubric questionnaire given for about 10% that is from 0.668 to 0.774.

5. Conclusion

This study aims to design and develop an aid that can give influence in helping dyslexic children solve their emotional problems by increasing their self-awareness. In line with this the implementation results show increasing in self-awareness assessment, this can be seen from the increases average score of rubric assessment of self-awareness for about 10%. It proves that children tend to improve themselves in looking for answers to things they have not mastered. It can be said that by using the CISELexia system they triggered to more curious about something. And the existence of the scoring system makes children have the motivation to improve themselves.

Overall this system still has possibilities for further improvement. Some suggestions that might be done for further research: add a variety of social story scenarios to increase data diversity in practicing *know your emotion* skills. It would be better to give leveling on learning emotion. Then it would be better to add a response made by the child against the social situation faced by the child to find out how far the *know your emotion* indicator is mastered.

References

[1] D. Bavelier, C. S. Green, and M. S. Seidenberg, "Cognitive development: gaming your way out of dyslexia?," *Curr. Biol.*, vol. 23, no. 7, pp. R282--R283, 2013.

[2] T. Cuschieri, R. Khaled, V. E. Farrugia, H. P. Martinez, and G. N. Yannakakis, "The iLearnRW game: support for students with Dyslexia in class and at home," in *2014 6th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES)*, 2014, pp. 1–2.

[3] S. M. Daud and H. Abas, "'Dyslexia Baca'Mobile App--The Learning Ecosystem for Dyslexic Children," in *2013 International Conference on Advanced Computer Science Applications and Technologies*, 2013, pp. 412–416.

[4] S. Franceschini, S. Gori, M. Ruffino, S. Viola, M. Molteni, and A. Facoetti, "Action video games make dyslexic children read better," *Curr. Biol.*, vol. 23, no. 6, pp. 462–466, 2013.

[5] M. H. Raskind and E. L. Higgins, "Speaking to read: The effects of speech recognition technology on the reading and spelling performance of children with learning disabilities," *Ann. Dyslexia*, vol. 49, no. 1, pp. 251–281, 1999.

[6] L. Rello, C. Bayarri, and A. Gorriz, "What is wrong with this word? Dyseggxia: a game for children with dyslexia," in *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*, 2012, pp. 219–220.

[7] L. Rello, C. Bayarri, Y. Otal, and M. Pielot, "A computer-based method to improve the spelling of children with dyslexia," in *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility*, 2014, pp. 153–160.

[8] M. R. U. Saputra and M. Risqi, "LexiPal: Design, implementation and evaluation of gamification on learning application for dyslexia," *Int. J. Comput. Appl.*, vol. 131, no. 7, pp. 37–43, 2015.

[9] J. P. Hourcade, N. E. Bullock-Rest, and T. E. Hansen, "Multitouch tablet applications and activities to enhance the social skills of children with autism spectrum disorders," *Pers. ubiquitous Comput.*, vol. 16, no. 2, pp. 157–168, 2012.

[10] A. M. Piper, E. O'Brien, M. R. Morris, and T. Winograd, "SIDES: a cooperative tabletop computer game for social skills development," in *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, 2006, pp. 1–10.

[11] F. D. D. Reed, S. R. Hyman, and J. M. Hirst, "Applications of technology to teach social skills to children with autism," *Res. Autism Spectr. Disord.*, vol. 5, no. 3, pp. 1003–1010, 2011.

[12] F. J. Sansosti and K. A. Powell-Smith, "Using computer-presented social stories and video models to increase the social communication skills of children with high-functioning autism spectrum disorders," *J. Posit. Behav. Interv.*, vol. 10, no. 3, pp. 162–178, 2008.

[13] A. Tartaro, J. Cassell, C. Ratz, J. Lira, and V. Nanclares-Nogués, "Accessing peer social interaction: using authorable virtual peer technology as a component of a group social skills intervention program," *ACM Trans. Access. Comput.*, vol. 6, no. 1, p. 2, 2015.

[14] J. S. K. H. E. Heiervang Anders Lund, "Behaviour problems in children with dyslexia," *Nord. J. Psychiatry*, vol. 55, no. 4, pp. 251–256, 2001.

[15] H. Tur-Kaspa, A. Weisel, and L. Segev, "Attributions for Feelings of Loneliness of Students with Learning Disabilities," *Learn. Disabil. Res. Pract.*, vol. 13, no. 2, pp. 89–94, 1998.

[16] J. D. Mayer, P. Salovey, P. Salovey, and D. Sluyter, "Emotional development and emotional intelligence: Implications for educators," *What is Emot. Intell.*, pp. 3–31, 1997.

[17] M. Narimani, A. Sadeghi, N. Homeily, and H. Siahpoosh, "A comparison of

emotional intelligence and behavior problems in dyslexic and non-dyslexic boys," *J. Appl. Sci.*, vol. 9, no. 7, pp. 1388–1390, 2009.
[18] D. Goleman, "Kecerdasan Emosional, terj," *T. Hermaya, Jakarta Gramedia Pustaka Utama*, 1996.
[19] D. Goleman, "Emotional intelligence. Why it can matter more than IQ.," *Learning*, vol. 24, no. 6, pp. 49–50, 1996.
[20] A. P. Association and others, "DSM: Diagnostic and Statistical Manual of Mental Disorders, VA: Arlington." 1994.
[21] A. All, E. P. N. Castellar, and J. Van Looy, "Assessing the effectiveness of digital game-based learning: Best practices," *Comput. Educ.*, vol. 92, pp. 90–103, 2016.
[22] P. Backlund and M. Hendrix, "Educational games-are they worth the effort? A literature survey of the effectiveness of serious games," in *2013 5th international conference on games and virtual worlds for serious applications (VS-GAMES)*, 2013, pp. 1–8.
[23] T. M. Connolly, E. A. Boyle, E. MacArthur, T. Hainey, and J. M. Boyle, "A systematic literature review of empirical evidence on computer games and serious games," *Comput. Educ.*, vol. 59, no. 2, pp. 661–686, 2012.
[24] A. Knutas, J. Ikonen, U. Nikula, and J. Porras, "Increasing collaborative communications in a programming course with gamification: a case study," in *Proceedings of the 15th International Conference on Computer Systems and Technologies*, 2014, pp. 370–377.
[25] E. Tsekleves, J. Cosmas, and A. Aggoun, "Benefits, barriers and guideline recommendations for the implementation of serious games in education for stakeholders and policymakers," *Br. J. Educ. Technol.*, vol. 47, no. 1, pp. 164–183, 2016.
[26] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: defining gamification," in *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, 2011, pp. 9–15.
[27] P. Di Bitonto, N. Corriero, E. Pesare, V. Rossano, and T. Roselli, "Training and learning in e-health using the gamification approach: the trainer interaction," in *International Conference on Universal Access in Human-Computer Interaction*, 2014, pp. 228–237.
[28] J. Hense, M. Klevers, M. Sailer, T. Horenburg, H. Mandl, and W. Günthner, "Using gamification to enhance staff motivation in logistics," in *International Simulation and Gaming Association Conference*, 2013, pp. 206–213.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 268-272 (2019)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Performances of Thermoelectric Cooling System for Cooling Flowing Water in a Water Block

Mirmanto Mirmanto*, Yesung Allo Padang, Zefri Ardian

Jurusan Teknik Mesin, Universitas Mataram, 83125, Indonesia

ARTICLE INFO

Article history:
Received: 02 August, 2019
Accepted: 14 September, 2019
Online: 08 October, 2019

Keywords:
Thermoelectric
Mass flow rate
Water block
COP

ABSTRACT

This study investigated thermoelectric cooling system performances to cool the water flowing inside a water block. The experiments employed two identical thermoelectrics model TEC2-25408 with an overall dimension of 40 mm x 40 mm x 6.4 mm. The cold side of the thermoelectric was affixed to the top surface of the water block, while the hot side of the thermoelectric was attached to the heat pipe. The mass flow rates of the water were varied, i.e. 1 g/s, 1.5 g/s, and 2 g/s. Calibrated K-type thermocouples with an uncertainty of ± 0.5°C were used for measuring the temperatures. The experiment was observed for about 5 hours. All data were recorded using the data logger DAQ MX 9714 NI that was connected to the PC with the LabView program. The results show that raising the mass stream rate increases the cooling capacity and the COP. However, the effect of the mass stream rate on the water temperature at the outlet is unclear. In general, the thermoelectric has less advantage for cooling flowing water.

1. Introduction

The cooling system for electronics develops continually in accordance with heat pipe technologies for removing heat rates from CPU. With its compact shape, high heat dissipation, and easy maintenance, the combination of heatsink-heat pipes becomes the right solution to cool the hot side of the thermoelectric module (TEM). The advantages of thermoelectric cooling system (TCS) are: (i) no heavy components, (ii) not noisy, (iii) small and compact, (iv) easy to maintain. However, the system also has disadvantages; e.g. small capacity, low COP (even less than 1) as reported in Ananta et al. [1], Mirmanto et al. [2, 3]. Ananta et al. [1] studied a thermoelectric cooler box for cooling drinking water. The aim of their study was to know the effect of thermoelectric electrical arrangements on the cooler box COP. The TEMs were arranged in parallel and series electrical connections. Meanwhile, Mirmanto et al. [2] investigated the effect of heat dissipation units on the cooler box performances. Nevertheless, they all used the TEMs to cool the water in plastic bottles placed inside the cooler box. The water did not flow at all. Cooling static water (not flowing water) seems to be easy because the water temperature decreases with time.

Different from cooling static water, cooling flowing water is more complex because the inlet water temperature is constant; it does not decrease with time. Therefore, this may need a lot of energy. When the thermoelectric used is just one or two, the water may not get low temperatures at the outlet. The temperature difference between the inlet and outlet is very small. This means that cooling purposes cannot be achieved. Azimi et al. [4] reported that the decreased temperature of flowing water that was cooled using thermoelectric was around 1.2°C. Ahamat et al. [5] cooled the flowing water using TEM with the hot side of the TEM cooled using water. They found their COP was greater than 1, i.e. 4.5. However, the water temperature at the outlet was not revealed. Recently, Ardian [6] conducted an investigation using TEM to cool flowing water inside a water block. He found a similar phenomenon. He obtained the difference in temperature between the inlet and outlet was just only 4°C. Further, studies on the use of TEMs, unfortunately, are still limited. Therefore, not much information can be reviewed.

In general, researches on TEMs that have been done are to cool stationary water; e.g drinking water, water in bottles, water in a tank, stationary air or solid materials. The flowing water or air is usually used to cool the hot side of the TEM; e.g. Jugsujinda et al. [7], Enescu and Spertino [8], Kin et al. [9]. For the above background, this study tries to explore the possibility of TEMs to cool flowing water. The aim of this study

*M. Mirmanto, Jl. Majapahit, no. 62, Mataram, +6282111738971, m.mirmanto@unram.ac.id

is to investigate wheater the TEM is proven able to reduce the temperature of streaming water significantly.

2. Research Method

2.1. Experimental Set-up

To know the ability of the TEM for cooling flowing water, experiments are set as follows. This study uses an experimental method, i.e. testing the research apparatus to obtain data that are required to answer or solve the problem statement. The schematic diagram of the research apparatus is shown in Figure 1. The apparatus consisted of a water block, two TEMs, two heat pipe units, a water pump, a small radiator, a flowmeter, and a small water tank. The material investigated was water flowing in a water block as indicated by Figure 2. The water block dimension was 200 mm x 40 mm x 10 mm, and it was made of aluminum. The water block comprised of 15 fins inside with a thickness of 1 mm. The mass flow rate of the water was varied; e.g. 1 g/s, 1.5 g/s and 2 g/s, and measured using a flowmeter model FL1000 with an accuracy of ± 0.1 g/s. Type K thermocouples, which were calibrated, were used to measure all temperatures and the total numbers of thermocouples employed were 18. The uncertainty of the temperature was ± 0.5°C obtained from calibration. The temperatures were recorded directly using a data logger NI-cDAQ9174 so that all temperatures were recorded simultaneously. Meanwhile, the power supplied to the TEMs was of approximately 82.2 W.

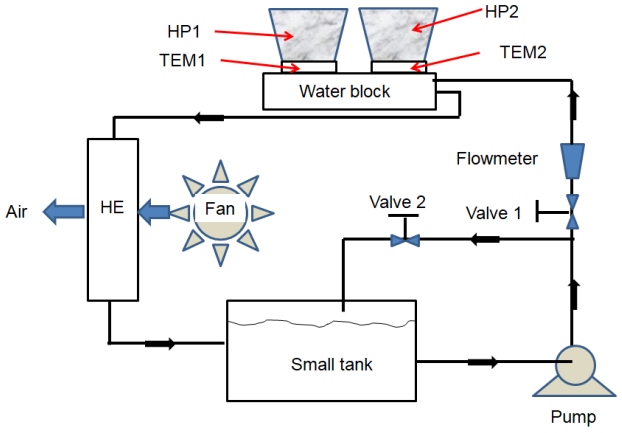


Figure 1: Schematic diagram of the research for cooling flowing water using TEMs, HP is heat pipe, HE is heat exchanger.

Two TEMs model TEC2-25408 were utilized in this study. They were installed on the water block. The cold side was faced to the water block, while the hot side was stuck on the heat pipe. Further, the water block was also insulated using an aluminum foil sponge. The heat rate from the water block was absorbed by the cold side of the TEMs, and then it was thrown away to the heat pipe on the hot side of the TEMs. The heat pipes employed were double fan heat pipe because each heat pipe had two fans. The photograph of the heat pipes installed and the dimension of the water block is shown in Figure 2.

Water was circulated throughout the test rig using a small pump. From the tank, water flowed to the flowmeter. Before passing through the flowmeter, the mass flow rate of water was adjusted using valves 1 and 2. Using this adjustment, the water mass flow rate could be controlled and suited. From the flowmeter, water then flowed to the water block, where two thermoelectrics were installed. The temperature of water decreased after passing the water block, and then water streamed to the heat exchanger. In the heat exchanger, the water gained heat again and the water temperature increased. From the heat exchanger, the water returned to the tank, and circulated again and so on.

2.2. Data Analysis

Analyzing experimental data can only be performed after some experiments are conducted. Data that are analyzed are temperatures, mass flow rates and power given to the TEMs. The data are analyzed using some equations that have been available in the published papers or books. From the analysis, some dependent variables are obtained and known; e.g. Q_c, Q_h, P, and COP. Then from these variables, a conclusion can be made. The conclusion will give us information about whether the TEM is potentially used to cool the flowing liquid or not.

In the use of Peltier principles, one side of the TEM becomes hot and the other becomes cool. The hot side temperature of the TEM can be estimated using equation (1) that can be found in Mirmanto et al. [3], Jugsujinda et al. [7], Abdul-Wahab et al. [10], Wahyu [11].

$$T_h = T_\infty + \theta Q_h \tag{1}$$

T_h is the hot side temperature of the TEM (°C), T_∞ represents the ambient temperature (°C), Θ is the thermal resistance of the TEM (°C/W) and Q_h is the heat rate (w) that is released by the hot side of the TEM.

The temperature difference between the cold side and the hot side of the TEM, ΔT, can be obtained using equation (2). This equation is adopted from [2, 3, 7, 10] and Mani [12].

$$\Delta T = T_h - T_c \tag{2}$$

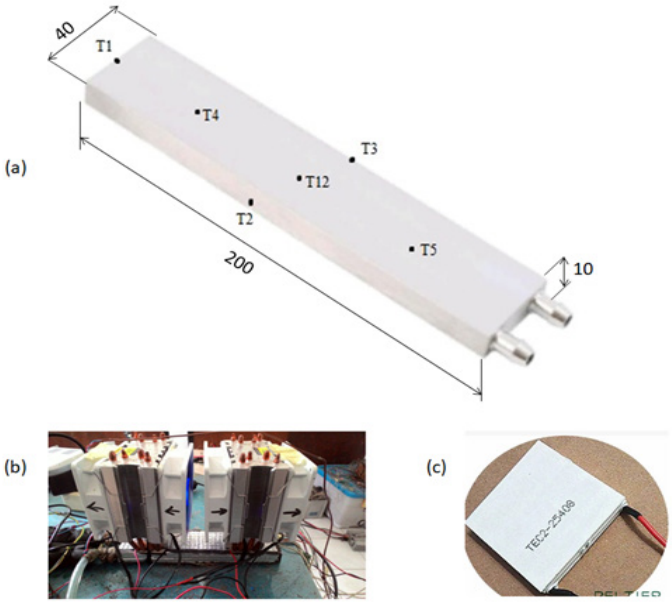


Figure 2: Water block and heat pipe; (a) positions of the thermocouples on the water block (unit in mm and without scaling), (b) double fan heat pipe installed on the water block, (c) TEM type TEC2-25408, Ardian [6].

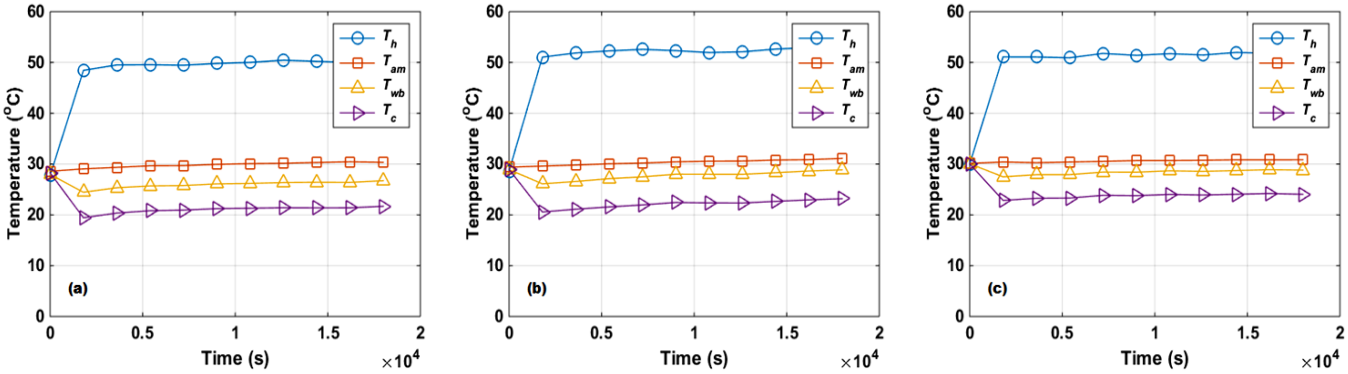


Figure 3: Relationship between temperatures and time; (a) 1 g/s, (b) 1,5 g/s, (c) 2 g/s. T_c: is cold side temperature of TEM, T_h: hot-side temperature of the TEM dan T_{wb}: average water block temperature.

$$Q_c = \dot{m} c_p \left(T_i - T_o\right) = \dot{m} c_p \Delta T_f \quad (4)$$

$\dot{m}$ is the water mass flow rate (kg/s), c_p is the specific heat (J/kg°C), and ΔT_f is the difference in temperature of the water, which is equal to the inlet temperature minus the outlet temperature at the water block (°C). c_p is determined using bulk temperature. The bulk temperature is equal to $(T_i + T_o)/2$ as suggested by Incropera et al. [16].

The electrical power supplied to the TEM can be estimated using equation (5), and it can be obtained in Incropera et al. [16], Cengel and Boles [17].

$$P = VI \quad (5)$$

T_h is the hot side temperature (°C) and T_c is the cold side temperature (°C). The relationship between heat rate of hot and cold sides of the TEM can be expressed by equation (3). This equation was taken from [1-3, 12].

$$Q_h = Q_c + P \quad (3)$$

Q_c is the heat rate of the cold side of the TEM (W), and P is the DC electrical power given to the TEM (W). Equation (3) was also used by Nugroho [13], and Ramdhan [14]. Q_c is also called a cooling load/ capacity that can be predicted using equation (4). This equation can be found in Holman [15], and Ardian [6].

P is the electrical power (W), V is the voltage (V), and I is the current (A). The voltage and current can be measured directly using a multitester. Meanwhile, the performance, which is noted by COP, can be predicted by equation (6), which was taken from Jugsujinda et al. [7]. The COP of the TEM usage is smaller than the COP of the compression refrigeration system as revealed in Wang et al. [18].

$$COP = \frac{Q_c}{P} \quad (6)$$

COP is the Coefficient of performance.

3. Results and Discussion

This study aims to cool water using a TEM with a heat pipe as the heat dissipation unit. The heat pipe is chosen because of the previous research has shown that the heat pipe can remove much heat rate compared to a heat sink with a fan as reported in Anggani [19]. The results of recording temperatures are shown in Figure 3.

Figure 3 indicates temperatures T_c and $T_{wb.}$ T_c and T_{wb} increased drastically at just after a second. After a few seconds, they become flat or even sometimes they increased. They increased because the room used for conducting experiments was not conditioned. When the ambient temperature increased, the temperatures of water and water block increased too. This actually indicated that the TEM could not absorb much heat from the water block containing flowing water. The TEM seemed not able to resist the small change in the ambient temperature. Meanwhile, the temperature of the hot side of the TEM increased drastically after a second and then it became constant. In Figure 3, the effect of mass flow rate on T_c is significant. The recorded T_c is presented in Table 1. Mass flow rate increases, *Tc* and *Twb* also increase. This indicates that at higher water mass flow rates, the TEM cannot decrease the outlet temperature at the water block exit. This phenomenon is really different from the thermoelectric cooling system performance for cooling stationary water. Cooling the stationary water could have lower temperatures, e.g. 13°C as reported in [2, 3], and Mirmanto et al. [20]. Furthermore, at around the initial time of running the machine, T_c and T_{wb} decreased sharply while T_h increased drastically. This was due to the character of the TEM. When the current flowed to the TEM, the cold and hot sides of the TEM suddenly changed. The cold side became cold abruptly, and the hot side became hot rapidly. However, after a moment, the heat transfer equilibrium took place, so that all temperature became constant or sometimes they raised due to the increase in ambient temperatures.

As shown in Table 1, T_c raised with an increase in mass flow rates. The temperature difference for each other was almost 1°C. This phenomenon was due to the constant power of the TEM. The power of the TEM affected greatly the capability of the TEM to absorb heat. Therefore, when the cooling load increased due to the rise in the mass flow rate, T_c elevated because the cold side of the TEM was no longer able to transfer heat. This was also found by Mirmanto et al. [20]. Nevertheless, T_c influenced the room and

water temperatures inside the cooler box. When T_c increased, the room and water temperatures raised as reported in [20]. In this study, when T_c elevated, the average water block also increased as shown in Figure 3.

Table 1: Recorded T_c at several mass flow rates

Time	T_c (°C)		
(s)	at 1 g/s	at 1.5 g/s	at 2 g/s
0	28.26	28.15	28.04
1800	19.43	20.57	22.85
3600	20.36	21.12	23.26
5400	20.82	21.56	23.31
7200	20.94	21.96	23.85
9000	21.23	22.44	23.78
10800	21.31	22.35	24.03
12600	21.41	22.34	23.91
14400	21.38	22.66	24.06
16200	21.39	22.95	24.20
18000	21.69	23.18	24.04

To analyze the ability of the TEM further, Figure 4 is presented. Figure 4 indicates the relationship between Q_c and time at several water mass flow rates. Figure 4 demonstrates that Q_c is constant with time. However, increasing the water mass flow rate elevates Q_c. This is suitable for the theory of Holman [15] or agrees with equation (4). The outlet temperature, *To*, decreased drastically at a few seconds, but then it increased gradually. This was due to the effect of ambient temperature. The research was started in the morning at 08.00 am and finished in the afternoon at 01.00 pm. Therefore, the ambient temperature increased, and consequently, the outlet water temperature, T_o, increased. Figure 5 shows the outlet water temperature, T_o. The effect of water mass flow rates on the outlet water temperature is unclear. The lines of outlet temperatures are touching each other at the uncertainty of ± 0.5°C. This could be due to the fluctuation of T_o. T_o and other temperatures were affected by the ambient temperature, therefore, T_o fluctuated depending on the ambient temperature.

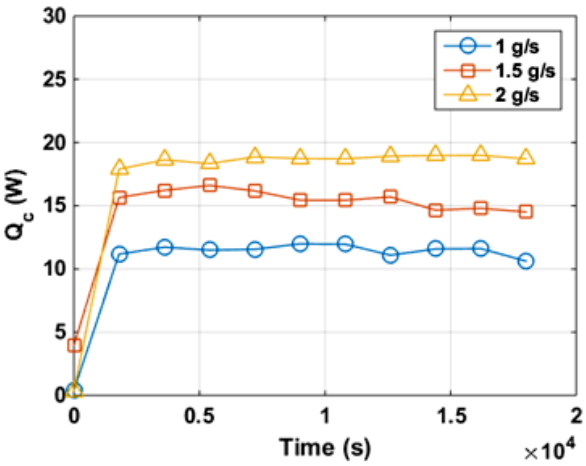


Figure 4: Relationship between Q_c and time at several mass flow rates of water

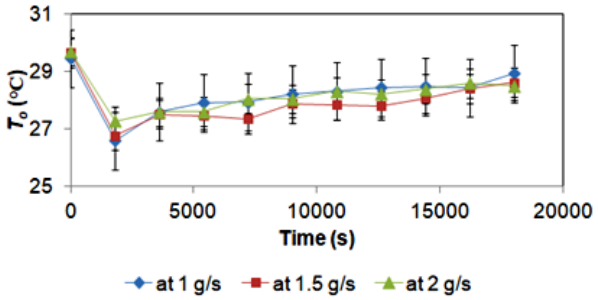


Figure 5: Relationship between *To* and time at several mass flow rates of water.

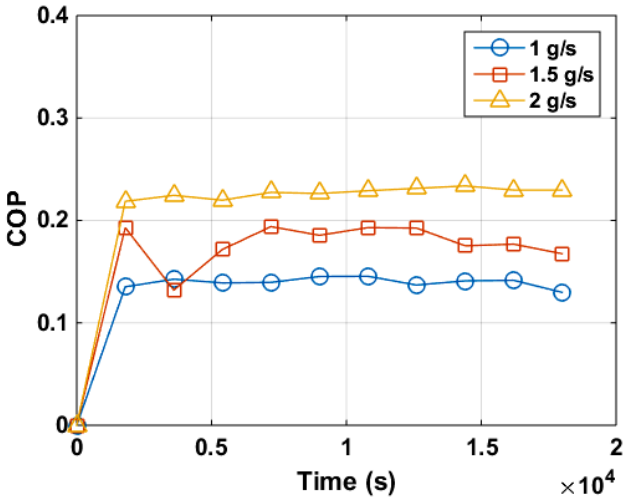


Figure 6. Relationship between COP and time at several mass flow rates of water

Another variable examined in this study is COP. The experimental COP is presented in Figure 6. COP indicates the performance of the cooler box, i.e. cooling capacity divided by power supplied to the TEM (or system) as expressed by equation (6). The power supplied to the TEM is calculated using equation (5). The calculated COP is given in Figure 6 at the three different mass flow rates. Similar to the cooling capacity's trend, COP increases drastically and then it gets persistent. This phenomenon occurs for the three mass flow rates. In Figure 6, it is clear that COP is affected by the mass flow rate. As the mass flow rate raises, COP elevates. This can happen due to increased cooling capacity, while the power supplied to the TEM is constant. This phenomenon was also found by Ramdhan [14], Mirmanto et al. [20]. Nevertheless, they used the cooler box for cooling water in plastic bottles. Their COP increased with an increase in the water volume. Further, COP obtained in this study and by Ramdhan [14] and Mirmanto et al. [20] is still lower than 1.

4. Conclusion

Experiments have been done to investigate the ability of the TEM to cool the flowing water. The TEMs used were two pieces with the electrical power of 82.2 W and the mass flow rates of the water employed were 1 g/s, 1.5 g/s, and 2 g/s. Some findings are as follows:

1. In general, the trend of the temperature in accordance with the results obtained by the previous researchers.
2. The effect of the mass flow rate on the cooling capacity

is clear. Increasing the mass flow rate elevates the cooling capacity.

3. The effect of the mass flow rate on the outlet temperature is unclear.
4. The temperature of the water block elevates as the mass flow rate increases. The minimum temperature of the cold side of the TEM at 1 g/s is 19.43°C, at 1.5 g/s is 20.57°C and at 2 g/s is 22.85°C.
5. COP raises with the increase in mass flow rate.
6. The water temperature is affected greatly by the ambient temperature.
7. Cooling flowing water using TEC is not powerful. The TEM is better to be used for cooling the stationary liquid.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

On this occasion, the authors would like to acknowledge the DRPM of the Indonesian Ministry of Research and Technology and Higher Education for the funding through the research scheme PDUPT 2019, with the contract number: **1842/UN18.L1/PP/2019**. Also, the authors thank the Department of Mechanical Engineering, Mataram University for the facility.

References

[1] H. Ananta, Mirmanto, Y.A. Padang, “Unjuk Kerja Kulkas Termoelektrik dengan Rangkaian Seri dan Parallel pada Beban Air 1500 ml” Dinamika Teknik Mesin, 7(2), 80-86, 2017. https://doi.org/10.29303/dtm.v7i2.157

[2] Mirmanto, I.B. Alit, I.M.A. Sayoga, R. Sutanto, Nurchayati, A. Mulyanto, “Experimental cooler box performance using two different heat removal units: a heat sink fin-fan, and a double fan heat pipe” Frontier in Heat and Mass Transfer, 10(34), 1-7, 2018. http://dx.doi.org/10.5098/hmt.10.34

[3] M. Mirmanto, S. Syahrul, Y. Wirdan, “Experimental performances of a thermoelectric cooler box with thermoelectric position variations” Engineering Science and Technology, an International Journal, 22, 177–184, 2019. https://doi.org/10.1016/j.jestch.2018.09.006

[4] M. Azimi, S.S. Mirjavadi, M. Sedighi, “A novel cooling system design for water block in liquid cooling garment” Journal of Science and Engineering, 07 (01), 072-082, 2016.

[5] M.A. Ahamat, R. Abidin, S.M. Abdullah, “Performance of thermoelectric module as a water cooler and water heater” International Journal on advanced science engineering Information Technology, 6(4), 524-528, 2016.

[6] Z. Ardian, “Sistem Pendinginan Coolant Menggunakan Termoelektrik dengan Unit Pembuang Panas Heat Pipe," Skripsi, Jurusan Teknik Mesin, Fakultas Teknik, Universitas Mataram, Mataram, NTB, 2018.

[7] S. Jugsujinda, A. Vora-ud, T. Seetawan, "Analysing of thermoelectric refrigerator performances" Procedia Engineering, 8, 154-159, 2011. https://doi.org/10.1016/j.proeng.2011.03.028

[8] D. Enescu, F. Spertino, "Applications of hybrid photovoltaic modules with thermoelectric cooling" Energy Procedia, 111, 904-913, 2017. https://doi.org/10.1016/j.egypro.2017.03.253

[9] L.K. Kin, A.T. Baheta, K. Habib, "Analytical investigation of thermoelectric performance for cooling application" Journal of Advanced Research in Fluid Mechanics and Thermal Sciences, 46(1), 32-40, 2018.

[10] S.A. Abdul-Wahab, A. Elkamel, A.M. Al-Damkhi, I.A. Al-Habsi, H.S. Al-Rubai'ey', A.K. Al-Battashi, A.R. Al-Tamimi, K.H. Al-Mamari, M.U. Chutani, "Design and experimental investigation of portable solar thermoelectric refrigerator" Renewable Energy, 34, 30-34, 2009. https://doi.org/10.1016/j.renene.2008.04.026

[11] D. Wahyu, "Pendingin Minuman Bercatu Daya Termoelektrik," Skripsi, Jurusan Teknik Mesin Politeknik Negeri Padang, Padang, 2016.

[12] P.I. Mani, "Design, Modeling and Simulation of a Thermoelectric Cooling System (TEC)," Mechanical and Aerospace Engineering Western Michigan University, USA, 2016.

[13] W. Nugroho, "Rancang Bangun Alat Pendingin Minuman Portable Menggunakan Peltier," Skripsi, Jurusan Teknik Mesin Universitas Muhammadiyah Pontianak, Kota Pontianak, 2017.

[14] G. Ramdhan, "Sistem Pendinginan pada Kulkas Termoelektrik dengan Variasi Pembebanan," Skripsi, Jurusan Teknik Mesin, Fakultas Teknik, Universitas Mataram, Mataram, NTB, Indonesia, 2016.

[15] J.P. Holman, Heat Transfer, 8th Edition, McGraw-Hill Companies, Inc., USA, 2002.

[16] F. Incropera, D.P. DeWitt, T.L. Bergman, A.S. Lavine, "Fundamentals of Heat and Mass Transfer, 6th Edition, John Wiley & Sons, Hoboken, 2006.

[17] Y.A. Cengel, M.A. Boles, "Thermodynamics an Engineering Approach", 6th Ed., McGraw-Hill Higher Education, Boston, 2008.

[18] Z. Wang, J. Zhang, Y. Liu, "A Practical Method for Measuring Thermal Conductance and Cooling Power of Thermoelectric Modules," Tinajing Institute Of Power Sources, China, 2001.

[19] Y. Anggani, “Peningkatan Peforma Kulkas Termoelektrik Dengan Menggunakan Heat Pipe Sebagai Unit Pembuang Panas," Skripsi, Jurusan Teknik Mesin Fakultas Teknik Universitas Mataram, Mataram, Indonesia, 2018.

[20] M. Mirmanto, I.W. Joniarta, I.M.A. Sayoga, N. Nurpatria, Y.A. Padang, I.G.N.K. Yudhyadi, "Effect of Water Volume on a Thermoelectric Cooler Box Performance" Frontiers in Heat and Mass Transfer, 11(17), 1-9, 2018. http://dx.doi.org/10.5098/hmt.11.17

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 273-280 (2019)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Proposal Methodology of Planning and Location of Loading/Unloading Spaces for Urban Freight Vehicle: A Case Study

Moufad Imane*, Jawab Fouad

Laboratory of Production, Energy & Sustainable Development (LPE2D), Sidi Mohamed Ben Abdellah University, Higher School of Technology (EST) Fez, Morocco.

ARTICLE INFO

Article history:
Received: 08 August, 2019
Accepted: 07 September, 2019
Online: 08 October, 2019

Keywords:
Urban logistics
Urban freight vehicles
Loading/unloading areas
Location
Planning

ABSTRACT

The problem of planning and location of loading/unloading (L/U) areas attracting interest from several researchers. This is mainly because of the heterogeneity of information and the great complexity of logistics flows. Among the projects of the urban logistics in Morocco, the L/U areas represent an essential importance in the deliveries to the centers of the agglomerations. In light of this context, this work proposes a methodology for planning and location of L/U spaces for urban freight vehicles taking a commercial street in Morocco as a case study. Therefore, the proposed framework details the different steps of planning of L/U areas since the elaboration of the commercial map of the study area until the phase of sizing and location. Then, we study the applicability of this system, making a detailed proposal of the physical arrangements will be put in place, and the principle of spatial and temporal organization of deliveries, aiming at the best match between available supply and demand.

1. Introduction

The urban freight transport (UFT) is an important activity that accounting for 15% to 20% of urban mobility [1, 2]. Thinking about UFT also means thinking about the organization of urban space, business development, the environment and people's quality of life. The arbitration between the different components of UFT and the uses of roads is particularly complex [3, 4].

Indeed, the road network and the public space represent a place of conflict between the UFT vehicles and other urban activities, particularly in mixed residential and commercial quarters. This conflict results in congestion, illegal parking and accidentality. Hence the better management of the organization associated with UFT vehicles is much needed. In this context, delivery areas represent a real opportunity in urban areas to facilitate the delivery/pick-up of UFT vehicles, ensure accessibility to delivery drivers, and reduce congestion [5].

In light of this context, this work proposes a methodology for planning and location of L/U areas in the case of a commercial street in Morocco. Our article begins with a state of the literature in order to understand the problem of organizing the transport of goods in the city in relation to the concept of L/U areas. The knowledge of theoretical and empirical work will give the efficacy to our work. The second step is to consolidate these achievements by proposing an approach in which we successively describe the different stages of preparation for the implementation of the L/U spaces. The next step will take the case of a commercial street in Morocco for the implementation of the proposed approach. We present the results obtained in this way.

2. Background

The terminology for loading/unloading areas has not standardized: delivery area, service area, or even, on some panels, stop minute can designate these spaces [6,7]. These spaces represent a stopping point for vehicles loading or unloading the goods [9]. They are located on public roads, usually in front of shops, supermarkets, restaurants or grocery stores. The term "stop" means the momentary stopping of a vehicle on a road for the time necessary to allow the loading or unloading of the freight vehicle [9].

There are many authors in the literature that focus on identifying best practices in the location and improvement of L/U spaces. In 2009, the French Center for Studies on Networks, Transportation, Urban Planning and Public Construction

*Moufad Imane, Email: imane.moufad@gmail.com

(CERTU [1]) published a guide on planning these spaces that provides a rough quantification of delivery area requirements, and a set of recommendations to improve their management.

In the same way, [7-9] proposed an approach to designing and operating the delivery areas stating the steps to achieve that.

In addition, [10] proposed the Monte-Carlo method to simulate the operation of L/U areas. We also refer to [11] who proposed the combination of two models (FRETURB & DALSIM) allowing the dimensioning and the allocation of the delivery zones by taking into account the impacts on the flows of global traffic and give an example of implementation in the city of La Rochelle.

In addition, [12] proposed a hybrid method between the FRETURB model and the CERTU method to quantify the demand for delivery areas. Even more, [13] proposed technical solutions to improve the number and locations of delivery spaces with a practical application on the city of Bologna in Italy.

The concept of delivery areas has also been studied by [8] who proposed a set of policy recommendations to increase the availability of these spaces by taking New York City as a case study. The authors [14] studied the concept of reservation and control of delivery areas in the case of a shopping street. They proposed a number of operating scenarios based on punctual deliveries and pickups, using a comprehensive database of UFT traffic and parking conditions.

Other relevant contributions to this area don't necessarily address all the design and layout practices of the L/U spaces but focus on the location and improvement of these spaces. In this sense, [15-17] assessed the impact of changing the number and location of delivery areas on the level of illegal parking practices. The authors in [18] developed for the case of the city of Seville, a model for optimizing the location of delivery spaces by evaluating the effect on the cost of the delivery system. In addition, [4] have developed a model for assessing and optimizing the location of delivery areas based on meeting the logistical needs of commercial establishments associated with the frequency of goods movements generated in the study area.

From the same angle, [19] addressed the problem of location of parking areas. They proposed a model to sample the arrival intervals and service times of trucks using these areas.

Finally, in another contribution, we find [20] who proposed, in their project Future Delivery Area, a new concept of operation of delivery areas based on the principle of prior booking by delivery drivers managed by a delivery information system. In addition, the same principle has been studied by [21] introducing the environmental parameter within the reservation of delivery areas.

According to this literature review, we conclude that existing research on the organization of urban freight transport across delivery areas is rich and can be divided into two parts. On the one hand, there is the theoretical and empirical work that has proposed methods for the design and operation of delivery areas in several stages. On the other hand, some works have not focused entirely on organizational practices but rather on the study of the impact of implementation or change of the configuration of delivery spaces on mobility, congestion and anarchic parking practices.

3. Proposed methodology

Indeed, a good management of the traffic and the parking of the vehicles of the UFT pass by a good management of the delivery areas in order to adapt the supply of areas to the demand of displacements [4]. To do this, it is necessary to know this demand exhaustively, which means: who? (What type of trade); When? (Critical time slots); What? (Type of vehicle that serves the trade); and how much? (Frequency of deliveries / pick-ups) [22,23].

Based on the theoretical and empirical work that we discussed previously, we propose an approach through which we address all the steps of the planning and the location of the delivery spaces in study area (Figure 1).

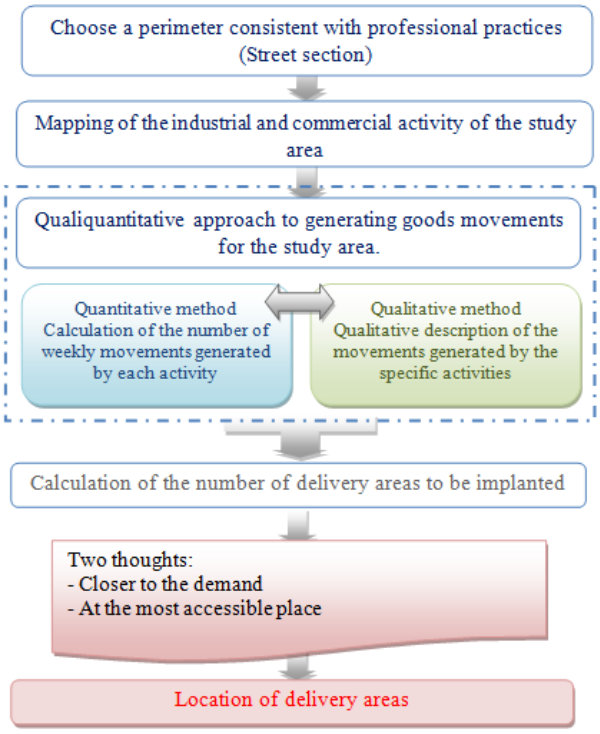


Figure 1: Proposed methodology of planning and location of L/U spaces.

3.1. Step 1: Choice of the study area

Although the delivery area development guide proposed by Certu[1] declares that there is really no minimum quota regulating the creation of delivery areas, we can remember that the implementation of a first delivery area can be done from 50 weekly movements [4-7]. Therefore, the study area must represent a living pond both functionally (employment, daily mobility) and economically (activity and product exchange basin) [21]. It should be in a densely populated area [3]. Given that the cost and time are required to carry out this type of study, it is

[1] CERTU Report, Planning delivery areas - Guide to quantification, their location and design, 2009, 52 pages.

necessary to proceed to a specification of the study area in order to guarantee a good representation of the commercial activity, which will serve as a basis for extrapolating the studies and hypotheses of movement generation at other study areas [6]. Thus, the area under study should allow us to observe, understand and measure how the UFT participates in the occupation of the road.

3.2. Step 2: Development of commercial and industrial mapping of the study area

This step consists of identifying the movements of all industrial and commercial activities in the study area [16]. Since it is a movement mapping, it must later allow a spatial presentation of the study area. By indicating the commercial typology on a background, it will be possible to identify traffic areas by activities, and thus give these zones characteristic movement generation ratios.

3.3. Step 3: Generation of movements

This step consists of determining the generation of goods movements through the census and the observation of the logistic functioning of all the industrial and commercial activities. According to [4,9,11,13,19] a movement generation study is based on two complementary methods: a quantitative method and a qualitative one. The prioritization of one method over the other is not specified in the above-mentioned works.

Quantitative method

This method is based on the criterion of movement ratio. From the commercial survey of the study area, it is necessary to define

an average coefficient of movements generated per week for each type of activity. Thus, it has been shown that the average number of movements generated by a point of sale is strongly correlating with the type of activity exerted by this point of sale [15,9,19]. Therefore, the number of delivery areas is calculated from the following relation[1]:

$$\text{Theoretical number of loading/unloading areas} = \frac{(\sum \text{Number of activities by type} \times \text{ratio of Mvts})}{(\sum \text{Mvts generated by a delivery area with full potential})} \quad (1)$$

*(Mvts: Mouvements)

Indeed, the knowledge of the movements generated per week and by typology makes it possible to establish the total number of movements generated in a given perimeter. Therefore, the theoretical number of delivery areas can be obtained by dividing this total by the average number of weekly movements that can be processed by L/U area used to its full potential. This theoretical number represents a first estimate of the need in the delivery area. The subsequent conduct of a field survey at the local level is necessary to refine the results obtained.

Qualitative method

This method is based on the results of field observations. It concerns the types of activities that are the subject of a qualification of their main characteristics as generators of movements of goods. Example of car dealers, banks ... etc.

In Table 1, we describe the characteristics and logistic functioning of this type of activity [6]: mapping of the movements generated at the level of the study area. However, as with many design principles, it is desirable to know the maximum stress to determine the provisions that will be implemented [9]. To refine this constraint, the motion analysis stage provides qualitative information on the nature of the movements generated according to the commercial typology [4].

Table 1 : Qualitative description of the movements of activity

Activity	Characteristic	Logistic operation
Coffee shops	Most of the time, this type of establishment receives daily delivery by specialized wholesalers. These long and bulky deliveries require proximity to the trade.	• One delivery per day • Long and bulky • Heavy weights
Department stores	Most of these establishments have integrated supply logistics (their own delivery areas and do not use those in the public space). However, some of them do not have private delivery areas.	• Several deliveries a day. • Integrated Supply Logistics. • Heavy weights
Car dealers	This type of establishment causes many problems during delivery such as street blocking. Most of the time, the delivery is done with an appointment.	• Very bulky • Delivery often taken with appointment.
Banks	This type of establishment requires specific treatment. In most cases, there is a reserved space for parking.	• Various movements per day • Requiring a specific parking space.
Pharmacies	This type of establishment is subject to frequent deliveries during the day, but these deliveries are generally short-term. It is therefore not always necessary to consider them, as the delivery areas are rarely used for such short stops.	• Frequent deliveries • Short term (2minutes)
Tertiary service	This activity class mainly generates small delivery or removal movements by courier companies. These movements happen on a daily basis and spread over the entire working period.	• Small movements. • Daily movements spread over the entire work period.

From these two complementary methods and information collected in the field, it is possible to draw up an effective

3.4. Step 4: Movements analysis

This stage of the diagnosis aims to determine the critical characteristics allowing the good dimensioning of the arrangements and the good logistic operating of the operations of deliveries / pick-ups in the study section. This is to describe the nature of the movements by type of business, specifying in particular the typology of vehicles and the handling tool. The necessary data for this step are [7-15]:

- The duration of the movements to define an average rotation rate and, apriority, the number of delivery areas required to ensure demand [24].
- The type of vehicles used to define the right-of-way of the delivery areas. The classification of vehicles can be done according to several criteria [4].
  - PTAC: This is the authorized loaded weight. It represents the maximum authorized mass for a road vehicle as stated in the Highway Code. It includes the weight of the unladen vehicle, the maximum load of goods and the maximum weight of the driver [25]. This criterion classifies the freight vehicles into two categories: light utility vehicles (LUVs) with a PTAC of less than 3.5 tones and heavy vehicles (HV) with PTAC greater than 3.5 tones.
  - Dimension: This includes the length and width of the vehicle as well as the floor area (length × width). These data are necessary for the size of the delivery areas.
- Handling tool used by the type of vehicle and the type of movement (pick-up or delivery). These data make it possible to approach the part of movements and vehicles requiring a handling tool. They provide an indicative image for the logistical diagnosis that serves as a guide for the layout of delivery areas in order to facilitate handling operations [21].

3.5. Step 5: Location of loading/unloading areas

This step is based on the analysis of accessibility [20]. From the traffic and parking conditions of goods transport vehicles, it is necessary to define the level of accessibility to the signs or service point. As a result, the analysis of accessibility is carried out on two levels: superficial analysis and specific analysis [4-18].

The first one indicates that the location of the loading/unloading area must be must determine in such a way that a parking management is sufficient to clear space for deliveries, and located in a way so as not to impede the visibility of users (professionals and particular) on other signaling equipment (bus area, signs, traffic lights ...) [11].

In the superficial analysis stage, the level of accessibility is roughly quantified at each point on the sidewalk, but without taking into account the differences between these points, that is to say [24]:

- How many movements do they generate?
- At what times of the day do they generate them?
- What types of vehicles serve them?
- Which handling tools use the freight vehicles?
- What is the time required for each of these deliveries?

4. Case study

4.1. Presentation of study area

Our choice was focused on the city of Fez as a case study. It is the second largest city in Morocco after Casablanca with a population of 1.1million in 2014[2]. The city accounts for nearly 62%[3] of the industrial fabric of the Fez-Meknes region, occupies a privileged place in the national industrial fabric and contributes to the socio-economic development of the region and consequently of the country.

As previously stated, the quantification of the need for loading/unloading areas must be carried out on a coherent perimeter representative of a high commercial density. Therefore, we chose the Mohammed Zerktouni Street as a case study. It is a single-store residential and commercial street (that is, one-way traffic) that stretches for 600 m. Located in the city center, the street begins with the entrance "Atlas" and ends with an exit to Avenue Mohammed Esslaoui. We divided this avenue into four sections. Considering the cost and the time of realization of this kind of study, we decided to carry out the survey on section 1.

The figure 2 presents the study area:

Figure 2: The study area –Mohammed Zerktouni Street

The study zone extends over a length of 205 m from the main entrance of Mohammed Zerktouni Street to the "Patrice LEMOMBA". Indeed, this part is characterized by the existence of two inputs and two outputs. At 95 m from the main entrance, there is a two-way road (Entrance / Exit) through which the vehicles of urban freight transport transit at the Mohammed Zerktouni Street. Thus, the width of the section is about 8 m; the municipality has reserved 2m on both sides for parking. This means that it is indeed a narrow commercial street.

[2] http://worldpopulationreview.com/countries/morocco-population/cities/

[3] http://fesmeknesinvest.ma/sites/default/files/Pr%C3%A9fecture%20de%20F%C3%A8s.pdf

In Figure 3, we have illustrated all the dimensions of section 1.

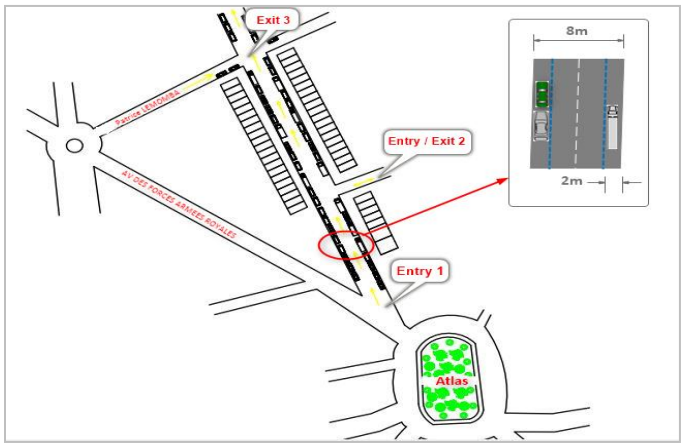


Figure 3: Study area layout

4.2. Mapping of commercial and industrial movements of the study area

To elaborate the mapping of the commercial and industrial movements of the study section, we conducted a survey that spanned a period from mid-November 2017 to the first week of April 2018. Through which, we will observe, understand and measure how the urban freight vehicles participates in the occupation of roads. This is a two-part inquiry. The first part concerns all the field visits carried out in order to understand the logistic operation, to identify the problems and to draw up the survey of the commercial and industrial typology of the study area.

Through the field visits, we developed the commercial and industrial mapping of the study area (Figure 4):

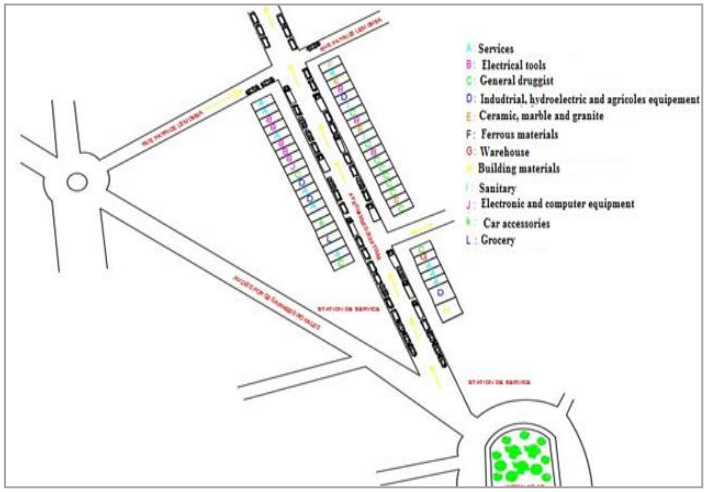


Figure 4: Mapping of commercial and industrial activity of the study area.

4.3. Quantification of the number of loading/unloading spaces

In this step, we have conduct a survey on the study area, in order to calculate the number of loading/unloading areas through the estimation of weekly movements generated by each establishment (Table 2).

According to the results of the survey, 60% of deliveries are making between 8:00 a.m. and 12:00 noon. Thus, a loading/unloading area can process an average of four deliveries per hour [22], bearing in mind that the commercial activity is

Table 2: The movements generated by industrial and commercial activities in the study area

Activity	Monday	Tuesday	Wednesday	Thursday	Friday	Sunday	Sum of movement per week	Number of service points	Average of movements generated per week / point of service	Mvts / point of service (Rounded to the next integer)
Services (Bank, Offices, Gas Station)	8	7	2	6	3	6	32	11	2,9	3
General drugstore	14	12	4	12	6	16	64	10	6,4	6
Warehouse	3	3	1	2	0	2	11	1	11	11
Industrial & hydroelectric equipment	7	5	3	6	2	8	31	4	7,8	8
Building Materials (Wood, Cement)	2	1	1	2	1	2	9	1	9	9
Ceramic, Marble & Granite	4	3	1	3	2	3	16	3	5,3	5
Grocery	2	2	1	2	2	2	11	1	11	11
Car accessories	1	1	0	1	1	1	5	1	5	5
Sanitairy	3	2	0	1	1	2	9	2	4,5	5
Electric tools	13	11	3	11	3	14	55	8	6,9	7
Ferrous materials	2	1	0	1	0	2	6	1	6	6
Automatic Electronics and Computer Equipment	2	1	0	2	0	3	8	1	8	8
TOTAL	61	49	16	49	21	61	257	TOTAL	83,8	84

carried out 6 times a week. Therefore, a delivery area can handle 6 * 16 deliveries, even 96 deliveries a week between 8h and 12h. Therefore, the number of loading/unloading spaces can be obtained by summing the weekly movements generated by each establishment (257 mouvements) and then dividing by the average number of weekly movements that can be treated by a loading/unloading area in its full potential. Therefore, the frame study zone requires the development of three loading/unloading areas as following:

Number of loading/unloading areas = 257/96 ≅ 3 (2)

However, it is important to know the conditions of movement generation to determine the provisions that should be implemented. The motion analysis part thus makes it possible to refine this constraint.

4.4. Movement Analysis

This step aims to determine the critical characteristics allowing the good dimensioning of the arrangements and the good logistic operating of the operations of deliveries / pick-ups in the study section. This is to describe the nature of the movements by type of business, specifying in particular the typology of vehicles and the handling tool.

We started with the typology of the vehicles borrowing the perimeter of study. From the field visits that we have made, we distinguish between four categories: Motorcycle & tricycle, cars, light utility vehicles (LUVs) and heavy vehicles (HV).

In figure 5, we present the results obtained from the survey about the use of these categories of vehicles according to the type of activities present in the study area:

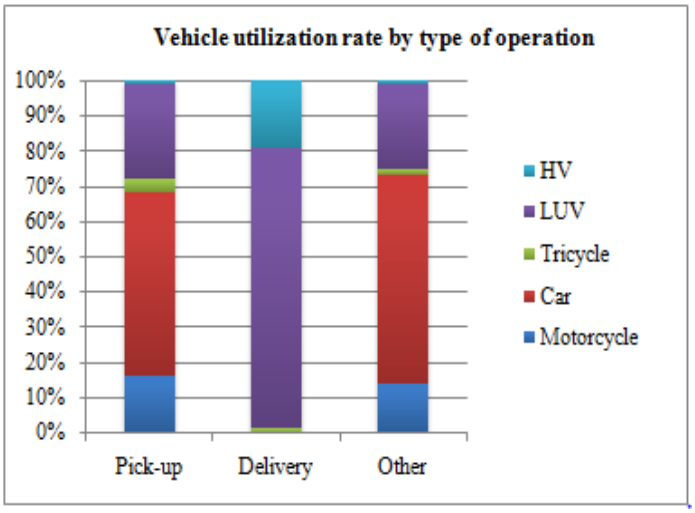


Figure 5: Distribution of mouvements by vehicle type.

The presence of heavy vehicles is a determining factor for the sizing of loading/unloading areas. The figure 5 show that 80% of deliveries operations are made by LUVs and 19% by trucks. The LUVs represents the category of vehicles with a length between 6 and 8 meters. The HV vehicles represents the category with a length between 8m and 15m. To ensure easy access to the area, the recommended dimensions are 12 to 15 meters in length and 2.20 to 2.50 meters in width for the size of the area [23].

The second point of analysis concerns the duration of the movements. It determines the possible turnover rate per loading/unloading area. The figure 6 shows the distribution of the duration of movements by type of movements:

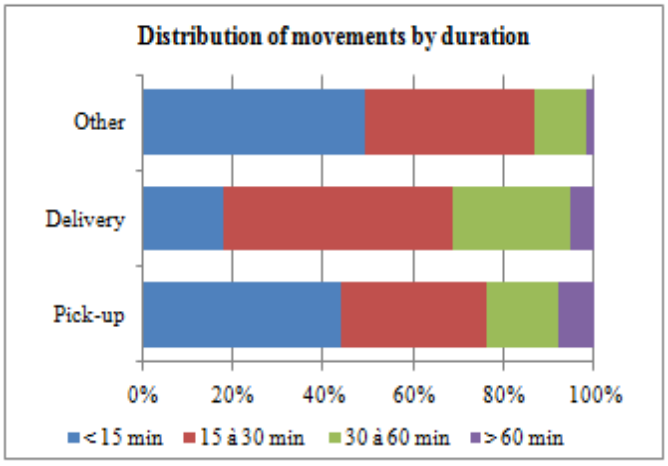


Figure 6: The distribution of the duration of movements by type of vehicle

By interpretation, more than 50% of pick-ups movements last more than 15 minutes. While 60% of delivery movements less than half an hour. The movements that last between 30 and 60 min represent mostly deliveries. For other types of operation, 90% do not exceed 30 minutes. Based on these results, the possible turnover rate per loading/unloading area is between 15min and 30 minutes.

Concerning the handling tools, the results of the survey show that 65% of the movements realized by the trucks lead to a devil, trolley and pallet truck that represents 30%. The figure 7 summarizes the results obtained:

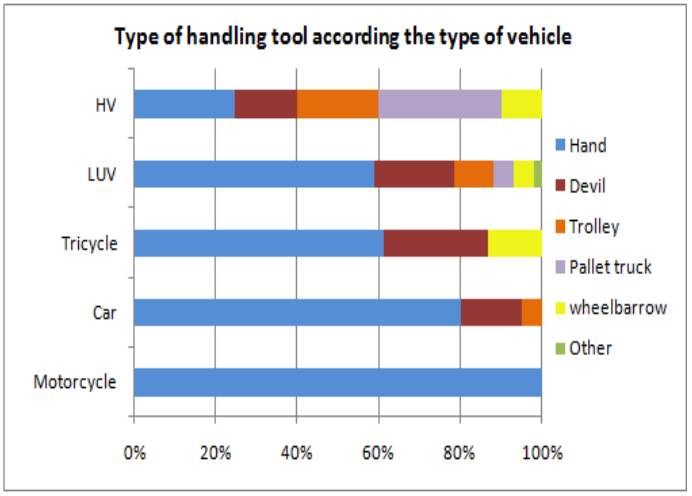


Figure 7: Use of handling tools by type of vehicle

For the LUVs, 60% of the movements are done by hand and 40% by devil, trolley, pallet truck and wheelbarrow. These data are essential for a good adaptation of the offer in delivery areas on demand.

For this, the lowering of the sidewalk for part of the area is important to facilitate handling operations.

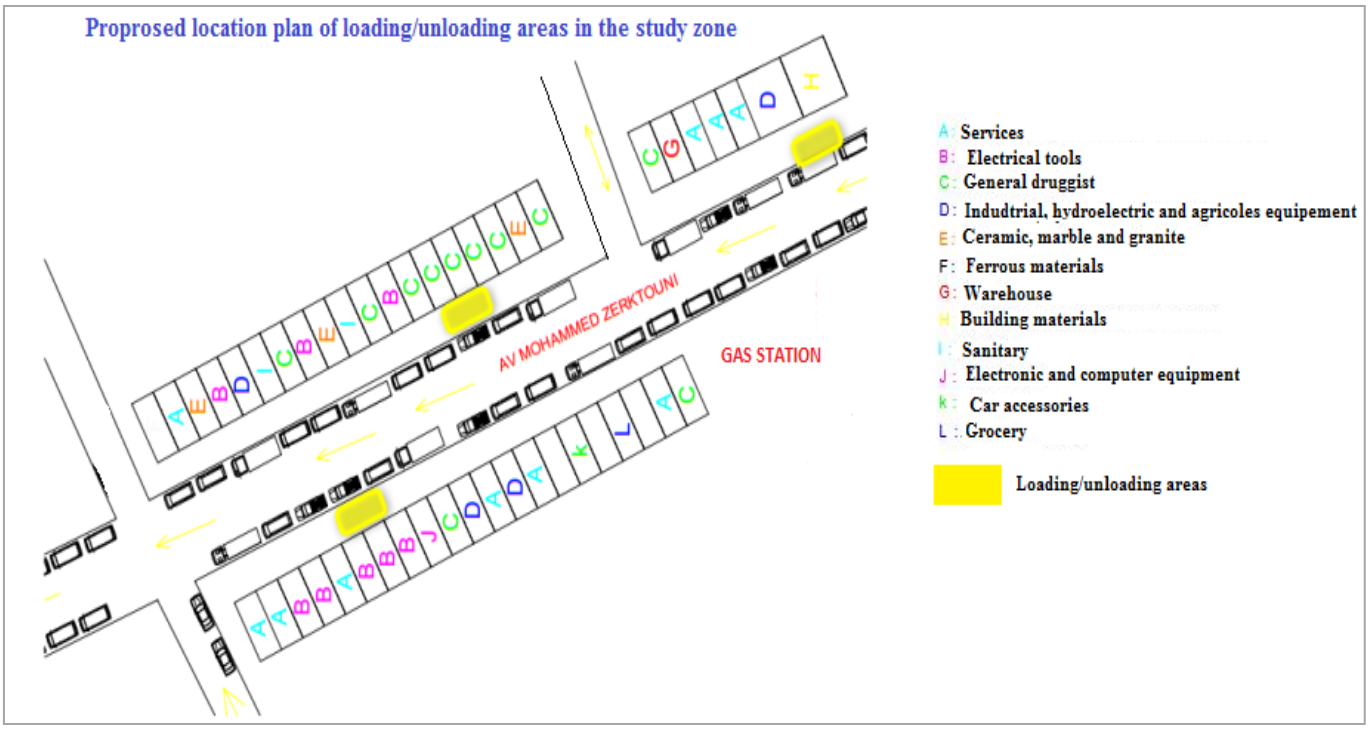


Figure 8: Location plan of loading/unloading areas in the study zone

4.5. Location of loading/unloading spaces

As mentioned earlier, the location of L/U areas is based on two types of analysis. On the one hand, the superficial analysis means that the location of the delivery area must be close to the shops receiving the highest frequency of deliveries and / or next to the activities giving rise to an important part of handling tools[4-18]. From the movement's analysis stage, we have an idea about the movements generated at each point of the study area. This information allows for a real assessment of the accessibility of driver-delivery drivers and a determination of areas likely to generate conflict.

In the case of the study area, this section represents the example of a narrow street with a strong single-line commercial activity lined by two parking lanes on both sides of 2m wide.

Therefore, it is interesting to drill a marked delivery area (that is to say on the sidewalk) since the left sidewalk is wide for (4m) and has no tree planting, and the other two straight half-notches (which means that is between the floor and sidewalk) since it measures 2.5m in width.

On the other hand, the specific analysis makes it possible to consider the typology of vehicles and handling tools in the location of L/U areas. Therefore, the location of the three loading/unloading areas must be close to the shops receiving the highest frequency of deliveries and / or next to the businesses giving rise to handling.

In our case, these are mainly: Building materials, general hardware, electrical tools, industrial, hydroelectric and agricultural equipment.

The figure 8 shows the location plan of the loading/unloading areas taking into consideration all of these elements:

5. Conclusion

In this article, we have proposed a methodology for planning and location of the loading/unloading spaces. Thus, we have presented all the works that dealt with the problem of design and location of these spaces. Based on this literature review, we have detailed all the steps of our methodology from the choice of the study area to the stage of location plan.

As a case study, we chose to deploy these steps in a commercial street in Morocco. It is the Avenue Mohammed Zerktouni of the city of Fez.

According to the results of the survey, 60% of deliveries are making between 8:00 a.m. and 12 noon, and the movements generated by each establishment are 257 mouvements. For this, the frame study zone requires the development of three loading/unloading areas.

In addition, we have described and presented in the movement's analysis step the nature of the movements by the type of vehicles, duration of movements and the handling tool in order to determine the critical characteristics allowing the good dimensioning and location of the L/U areas in the study section. Therefore, the location of the three L/U areas must be close to the activities receiving the highest frequency of deliveries and / or next to the activities giving rise to handling. In our case, these are mainly: Building materials, general hardware, electrical tools, industrial, hydroelectric and agricultural equipment.

The main contribution of this work is to propose an approach that decision-makers (especially local authorities) can easily apply to capture urban needs quickly and efficiently and thus allow for simple replication of the proposed framework in other cities.

As a perspective, an assessment of the impact of the location of loading/unloading areas on traffic flow will be the subject of future contributions. The objective is to evaluate how much L/U activities on the road affect traffic conditions that cause traffic congestion.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] L. Dablanc, G. Giuliano, K. Holliday, T. O'Brien "Best Practices in Urban Freight Management: Lessons from an International Survey" Transportation Research Record, 2379, 29–38, 2013. https://doi.org/10.3141/2379-04

[2] I. Moufad, F. Jawab, "A study framework for assessing the performance of the urban freight transport based on PLS approach" Archives of Transport, 49(1), 87-98, 2019. https://doi.org/10.5604/01.3001.0012.8368

[3] I. Moufad, F. Jawab, "Conception and Validation of a Decision Support Model for Urban Freight Transport" In IEEE International Colloquium on Logistics and Supply Chain Management (LOGISTIQUA), 94-99, 2017. https://doi.org/10.1109/LOGISTIQUA.2017.7962880

[4] R. Pinto, R. Golini, A. Lagorio, J. Gonzalez-Feliu, C. Ambrosini, P. Pluvinet, M. Serouge, "Loading/unloading lay-by areas location and sizing: a mixed analytic-Monte Carlo simulation approach", IFAC-Papers Online, 49(12), 961–966, 2016. https://doi.org/10.1016/j.ifacol.2016.07.900

[5] I. Moufad, F. Jawab, "The Determinants of the performance of the Urban Freight Transport - An Empirical Analysis" In IEEE International Colloquium on Logistics and Supply Chain Management (LOGISTIQUA), 2018. https://doi.org/10.1109/LOGISTIQUA.2018.8428296.

[6] M. Laurence, "The delivery area: Its use, its users, the principles of development". CERTU, 1-5, 2013.

[7] J. L., Reynaud, "Design and operation of delivery areas", 2015.

[8] M. Browne, J. Allen, A. Woodburn, D. Patier, J.-L. Routhier, C. Ambrosini, M. Jaller, "Comprehensive Freight demand Data Collection Framework for Large Urban Areas" Journal of Chemical Information and Modeling, 1-13, 2014. https://doi.org/10.1007/978-3-642-31788-0_6

[9] W. Debauche, "Delivery areas: for what, how much, where, how ?" In Centre de recherche routières (Ed.), INFRA. Quebec, 2015.

[10] A. Gaudron, S. Tamayo, A. De la Fortelle, "Simulation de l'utilisation des aires de livraisons en ville par la méthode de Monte-Carlo" In Les rencontres de la mobilité intelligente, 24-25 January, Paris, France, 2018.

[11] L. Delaître and J. L. Routhier, "Mixing two French tools for delivery areas scheme decision making" Procedia - Social and Behavioral Sciences, 2(3), 6274–6285, 2010. https://doi.org/10.1016/j.sbspro.2010.04.037

[12] M. Gardrat and M. Serouge, "Modeling Delivery Spaces Schemes: Is the Space Properly used in Cities Regarding Delivery Practices?" Transportation Research Procedia, 12, 436–449, 2016. https://doi.org/10.1016/j.trpro.2016.02.077

[13] G. Dezi, G. Dondi, C. Sangiorgi, "Urban freight transport in Bologna: Planning commercial vehicle loading/unloading zones" Procedia - Social and Behavioral Sciences, 2(3), 5990-6001, 2010. https://doi.org/10.1016/j.sbspro.2010.04.013

[14] F. McLeod and T. Cherrett, "Loading bay booking and control for urban freight", International Journal of Logistics Research and Applications, 14(6), 385-397, 2011. https://doi.org/10.1080/13675567.2011.641525

[15] A. R. Alho, J. de Abreu e Silva, de Sousa, J. P., E. Blanco, "Improving mobility by optimizing the number, location and usage of loading/unloading bays for urban freight vehicles" Transportation Research Part D: Transport and Environment, 61, 3-18, 2016. https://doi.org/10.1016/j.trd.2017.05.014

[16] A. Alho, J. de Abreu e Silva, J. P. de Sousa, "A state-of-the-art modeling framework to improve congestion by changing the configuration/enforcement of urban logistics loading/unloading bays" In 16th Meeting of the EURO Working Group on Transportation, Procedia - Social and Behavioral Sciences 111, 360-369, 2014. Doi: 10.1016/j.sbspro.2014.01.069

[17] A. Comia, B. Buttarazzi, M.M. Schiraldi, R. Innarella, M. Varisco, L. Rosati, "DynaLOAD: a simulation framework for planning, managing and controlling urban delivery bays" In the 19th EURO Working Group on Transportation Meeting, Istanbul, Turkey, Transportation Research Procedia, 22, 335–344, 2017. 10.1016/j.trpro.2017.03.049

[18] J. Muñuzuri, P. Cortés, R. Grosso, J. Guadix, "Selecting the location of minihubs for freight delivery in congested downtown areas" Journal of Computational Science, 3(4), 228–237, 2012. https://doi.org/10.1016/j.jocs.2011.12.002

[19] R. Pinto, R. Golini, A. Lagorio, J. Gonzalez-Feliu, C. Ambrosini, P. Pluvinet, M. Serouge, "Loading/unloading lay-by areas location and sizing: a mixed analytic-Monte Carlo simulation approach", IFAC-Papers Online, 49(12), 961–966, 2016. https://doi.org/10.1016/j.ifacol.2016.07.900

[20] D. Patier, B. David, R. Chalon, & V. Deslandres, "A New Concept for Urban Logistics Delivery Area Booking." Procedia - Social and Behavioral Sciences, 125, 99-110, 2014. https://doi.org/10.1016/j.sbspro.2014.01.1459

[21] L. Ramoneda Cuenca, "The problem of the last kilometer in Paris intra-muros. Introduction of the environmental parameter in the reservation of delivery areas". Polytechnic University of Catalunya, 2014.

[22] T. Baudel, L. Dablanc, P. Alguiar-Melgarejo, J. Ashton, "Optimizing Urban Freight Deliveries: From Designing and Testing a Prototype System to Addressing Real Life Challenges" Transportation Research Procedia, 12, 170–180, 2016. https://doi.org/10.1016/j.trpro.2016.02.056

[23] N. Aiura, E. Taniguchi, Planning on-street loading-unloading spaces considering the behaviour of pickup-delivery vehicles. In R. G. Thompson, & E. Taniguchi (Eds.), Recent advances in city logistics Oxford: Elsevier, 2006.

[24] P. Barette, M. Thoulen, E. Caelen, "Organization of road deliveries" The Mobility and Road Safety Monitor's notebooks, 8, September, 2012.

[25] Piarc. "Guide for setting up a road freight management system in an urban environment", 2012.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 281-286 (2019)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Sentiment Analysis of Transjakarta Based on Twitter using Convolutional Neural Network

Kevin Yudi*, Suharjito

Binus Graduate Program – Master of Computer Science, Computer Science Department, Bina Nusantara University, 11480, Indonesia

ARTICLE INFO

Article history:
Received: 29 August, 2019
Accepted: 27 September, 2019
Online: 08 October, 2019

Keywords:
Sentiment Analysis
Deep Learning
CNN

ABSTRACT

TransJakarta is one of the methods to reduce congestion in Jakarta. However, the number of TransJakarta users compared to number of private vehicle users is very small, only 24% of the total population in Jakarta. The purpose of this research is to know public opinions about TransJakarta whether positive or negative by doing sentiment analysis about TransJakarta based on the opinion of Twitter, as Twitter is one of media to express its many users to express their opinions about an individual or an instance. Data is retrieved from Twitter using the R-Studio application by utilizing the "TwitteR" library, then pre-processing and stored in a database. Next step is labelling the data using Sengon Lexicon and will be trained and tested using the Convolutional Neural Network algorithm. There are three CNN architectural models to be tested, namely VGG, ResNet, and GoogleNet. The designed VGG consists of 16 layers, ResNet 34 layers, and GoogleNet 22 layers. After the data are trained and tested, the results will be evaluated using Confusion Matrix to get the best F-Score. The results showed that among the three architectural models that were tested, the Resnet 34 layers architecture model gave the best F-Score of 98.11%, better compared to VGG which had the highest F-Score value of 96.74% and GoogleNet of 96.80%.

1. Introduction

Social Media is a popular platform to share information such as daily live updates, opinion and emotion in form of including pictures, text, videos, and audio. In Indonesia, the number of Social Media users is 90% of Internet users, counting to 132 million Users [1,2]. One of the platforms that are widely used is Twitter, where the number of users is 27% of the number of social media users in Indonesia. With those numbers of Users, Twitter can be used to retrieve information needed to do Sentiment Analysis. Sentiment analysis or opinion mining is a process of understanding, extracting and processing textual data automatically to get sentiment information contained in an opinion sentence [3].

One interesting phenomenon is the level of congestion in Jakarta. Based on the Inrix, Jakarta is ranked 12th as the most congested city in the world, and the second is in Asia. The government itself has sought to overcome severe congestion in Jakarta, one of which is by providing TransJakarta facilities. TransJakarta is a city-wide bus service that operates every day, with service coverage reaching all of Jakarta. The number of buses available by the end of 2017 has reached 3000 units, but TransJakarta's average number of users per day is only 340,000 people. By using Social Media, opinion about TransJakarta can be retrieved, and then can be used to analyze public image about TransJakarta. In the process of getting results, the right and accurate methods are needed [4]. One method that can be used is using Deep Neural Network Convolutional Learning.

In-depth learning in the last decade achieved satisfactory results in image analysis and analyzing speech in the form of text. One of the developing models is the Convolutional Neural Network (CNN). CNN is a model of artificial neural network that does not use the steps carried out by traditional artificial neural networks, but uses the convolution method while to produce output, input data will go through many different filters, then the results of this filter will be combined so that the results obtained more accurate [5]. Some CNN architectures are VGG, ResNet, and GoogleNet, each of which has a different number of layers. In this study, CNN Deep Learning will be used to obtain the results of social media data analysis about TransJakarta, then it will seek the best verification by comparing existing CNN architectural models.

* Kevin Yudi, Email: kevin.yudi@binus.ac.id

2. Related Works

In 2014, Dos Santos and Gatti [6] conducted sentiment analysis of data from Stanford Sentiment Treebank (SSTb), which contained sentences from the results of the film's connection, Stanford Twitter Sentiment Corpus (STS), which contained text from Twitter. This research proposes the Character to Sentence Convolutional Neural Network (CharSCNN) method, using two convolution layers to extract features related to words and sentences in various sizes. From this experiment, Dos Santos and Gatti produced an AccuracyI of 86.4%. This research has not used MaxPooling to measure the size of the output produced.

Then in 2015, Severyn and Moschitti [7] conducted Twitter data sentiment analysis using the Dynamic Nevolute Neural Network. Research from Severyn and Moschitti emphasizes the use of Word Embedding, between Word2Vec and Random Word Embeddings in Information Retrieval by using an unsupervised neural language model to train initial word embeddings that are further tuned by deep learning model on a distant supervised corpus. This study resulted in an accuracy of 87.12%.

In 2017, Yenter and Verma [8] conducted a sentiment analysis using Film Review Data from the IMDB website using Deep CNN-LSTM with a Combined Kernel of Various Branches for Analysis of IMDb Review Sentiments. This research tries to try one-dimension kernel with size 3, 5.7 and 9. Each branch's LSTM layer has 128 units. Any less or more units reduce accuracy or increase overfitting. From this research it produces an Accuracy of 89.5%.

In 2018, Cano and Morisio [9] also conducted sentiment analysis using Film Review Data from the IMDB website. The architectural model used is NGramCNN, compared to the SingleCNN and BLSTM-2DCNN models. It uses pretrained word embeddings for dense feature representation and a very simple single-layer classifier. The classifier used consists of a dense layer of 100 units and L2 regularization with 0.09 weight, followed by the output layer. Dropout of 0.5 between the dense and output layers were also used to avoid overfitting. From this study resulted in an accuracy of 91.2%.

3. Methodology

The first step taken in research is to retrieve data from Twitter. Data taken from Twitter contains the keyword 'TransJakarta'. Twitter data can be obtained using the API provided by Twitter [10]. After receiving data about TransJakarta, the data will then be stored in a database. The database used is PostgreSQL version 10.

After the data is stored in a database, the data will be processed and stored so that the data can be used to analyze sentiments. This process is called PreProcessing [11-13]. After completion, the data is then labeled per Tweet using a dictionary or Lexicon. The Lexicon used is the Lexon Sengon. After that the data will be entered in CSV and will be used as a data model for the Train and Test of the Learning algorithm in using several CNN models compared to one RNN model.

The CNN model used consisted of VGG consisting of 16 layers [14], Residual Net (ResNet) consisting of 34 layers [15], GoogleNet consisting of 22 layers [16]. While the RNN model used is LSTM. The experiment will be carried out in two variations, namely using Train 80 data compared to Test 20 and Train 90 data compared to Test 10. After the Training Model data model is formed and generate Test data, then the data will be evaluated using the Confusion Matrix. Figure 1 shows the steps of this research.

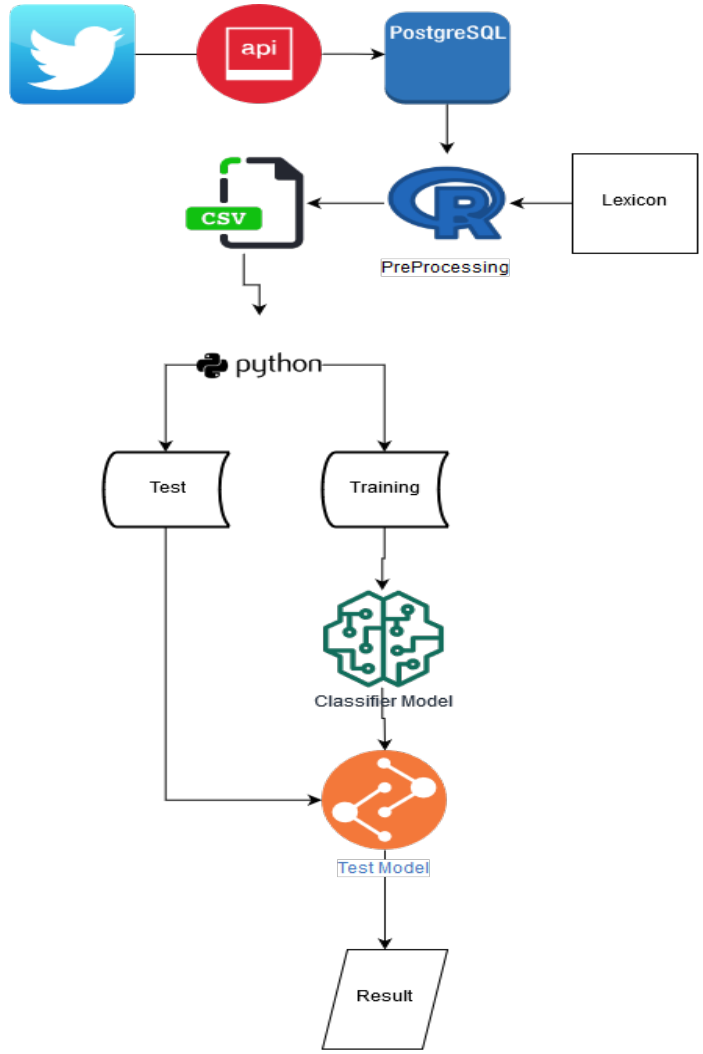


Figure 1: Methodology

3.1. Data Crawling

Data from Twitter will be collected through the Twitter API using R-Studio software. The data collected are about TransJakarta referred by Twitter users from August 1, 2018 to October 31, 2018. After the data is obtained, the data is then stored in the PostgreSQL Database. But not all parts of the data are stored in a database. Information needed for research is as follows:

o TweetID: The ID for each Tweet that exists, originating from Twitter.
o TweetContent: Tweet from Twitter about TransJakarta.
o TweetDate: The date a Tweet was posted.
o InsertDate: Date the data was inserted into the database.

3.2. PreProcessing

PreProcessing is a step to clean up data for sentiment analysis. The PreProcessing steps are as follows:

3.2.1 Case Folding

This step aims to turn all words into lowercase letters. The aim is to avoid case sensitive when matching words with a dictionary. An example is the change in the word 'Slow' to 'slow'.

3.2.2 Normalization

This step aims to remove the link in the post on social media, because the link is not part of the analysis. For example, the 'bit.ly/Xoa81p' link will be removed as ''. This step is also carried out with the aim of avoiding promotional spam using unclear links.

3.2.3 Data Cleansing

Data cleansing is the process of removing characters other than letters, such as punctuation and symbols. In this process also removes the 'RT' symbol which indicates that this Tweet is a Tweet from another user's Tweet.

3.2.4 Removing Stopwords

This step aims to eliminate words that are considered to have no meaning. For example, the words 'there' and 'what' are deleted because they have no meaning.

3.2.5 Tokenization

This step is done by separating each word into one separate part. Separation of these words is done by cutting sentences based on spaces so that later can be made a vocabulary based on unique words contained in the text.

3.3. Labelling Process

1. Enter the Punjabi text as input.
2. Divide this Punjabi paragraph into tokens and store the words in an array list.
3. Select the first word from array list.
4. Fetch the words of database in second array named as database array.
5. Check whether selected paragraph word matched with each word of database array.

(i) If match found

(a) Find the sentiment of word from database whether it is positive/negative or neutral.

(b) Find the exact position of word in the paragraph.

(c) Highlight the word according to their sentiment; make it green if it is positive, red if it is negative and blue if it is neutral.

(d) Calculate the score of sentence.

(e) Store the results in database.

(ii) Else match not found

(a) Select next word from the array

(b) Go to step 5.

6. Display the result to the user.
7. Plot the graph according to the results.

Labeling Process is a step to give a positive or negative label to a Tweet based on the words contained in the Tweet. This step is done by comparing the words in the Tweet with the list of words contained in Lexicon, both for positive and negative categories [17,18]. Figure 2 shows the algorithm of Lexicon based Sentiment Analysis. The Lexicon used for this study is Sengon Lexicon, which can be obtained at the following link: https://github.com/masasdani/sengon. Sengon Lexicon consists of 3061 positive words and 4239 negative words.

3.4. Convolutional Neural Network

Convolutional Neural Network is one of the machine learning methods of developing Multi-Layer Perceptron (MLP) which is designed to process two-dimensional data but can be used for text classification [19-21]. CNN is included in the Deep Neural Network type because of its deep network level and is widely implemented in image data. CNN has two methods; namely classification using feedforward and learning stages using backpropagation. The way CNN works is similar to MLP, but in CNN each neuron is presented in two dimensions, unlike MLP where each neuron is only one dimensional. In purely mathematical terms, convolution is a function derived from two given functions by integration which expresses how the shape of one is modified by the other.

Several architectural models will be tested in this study, including VGG, ResNet, and GoogleNet. The three models will be designed using the same hyperparameter, but with a number of different layers according to the character of each model. Table 1 shows the paramaters that used in this research.

Table 1 – Comparison of Proposed Model Architecture

	VGG	**ResNet**	**GoogleNet**
Layers	16	34	22
Epoch	5	5	5
Batch Size	32	32	32
Activation Function	ReLU	ReLU	ReLU
Max Pooling	50%	50%	50%
DropOut	0.5	0.5	0.5
Kernel Size	3x3	3x3	3x3

The first experiment will be carried out using Training Data 80% compared to Test Data 20%. The second experiment will be carried out using Training Data 90% and Test Data 10%. Each of Training data will be evaluated using *k*-fold cross validation method. In k-fold cross-validation, the original sample is randomly partitioned into k equal sized subsamples. Of the k subsamples, a single subsample is retained as the validation data for testing the model, and the remaining k − 1 subsamples are used as training data [22]. The cross-validation process is then repeated k times, with each of the k subsamples used exactly once as the validation data. The k results can then be averaged to produce a single estimation. The advantage of this method over repeated random sub-sampling is that all observations are used for both training and validation, and each observation is used for validation exactly once. In this research, number of the fold $k = 5$.

$$((f * g)(t) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau) \qquad (1)$$

3.5 Evaluation Method

Evaluation is a step to get the accuracy value from the model that has been made. The evaluation method used is Confusion Matrix. Confusion Matrix is used to get values consisting of Accuracy, Precision, Recall, and F-Score [23].

Accuracy: The percentage of data that is correctly identified is compared with the sum of all data.

$$\frac{TP + TN}{TP + TN + FP + FN} \qquad (2)$$

Precision: The total number of correctly classified positive examples by the total number of predicted positive examples.

$$\frac{TP}{TP + FP} \qquad (3)$$

Recall: The ratio of the total number of correctly classified positive examples divide to the total number of positive examples.

$$\frac{TP}{TP + FN} \quad (4)$$

F-Score: The harmonic mean of the *precision* and *recall*, where an *F-score* reaches its best value at 1 (perfect *precision* and *recall*) and worst at 0.

$$\frac{2(Precision * Recall)}{(Precision + Recall)} \quad (5)$$

Definition of terms:
True Positive (TP): Observation is positive and is predicted to be positive.
False Negative (FN): Observation is positive but is predicted negative.
True Negative (TN): Observation is negative and is predicted to be negative.
False Positive (FP): Observation is negative but is predicted positive.

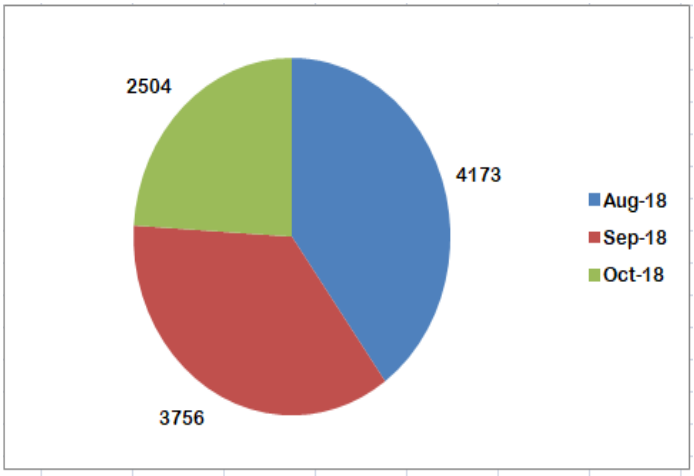


Figure 3: Chart of Data Distribution

4. Results Analysis

4.1. Data Collection

Data for this research were taken in the period between August 1, 2018 and October 31, 2018. In August 2018, the number of Tweets obtained was 4173 Tweets. While in September 2018 passed 3756 Tweets. Last October 2018 exceeded 2504 Tweets, with a total data of 10433 Tweets. Due to the policy of the Twitter API which only allows retrieving data from the past week, the data is taken every two days within a week. Figure 3 shows the distribution of data taken for this research.

From the results of the labeling process, we got Tweet data labeled 'Positive' of 7174 Tweets out of a total of 10433 Tweets or 68.76%. Whereas the Tweet labeled 'Negative' by 3259 or 31.24%. Data labeled Neutral were not used in this research because it did not provide any meaning between positive and negative. The percentage distribution can be seen in Figure 4.

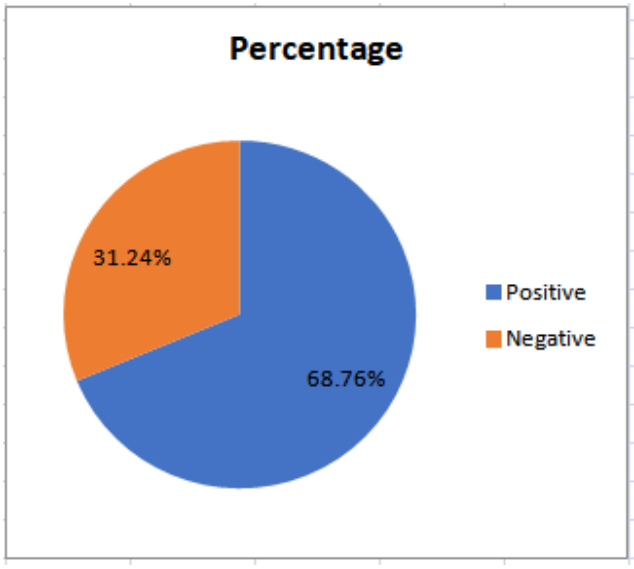


Figure 4: Percentage of Data after Labelling

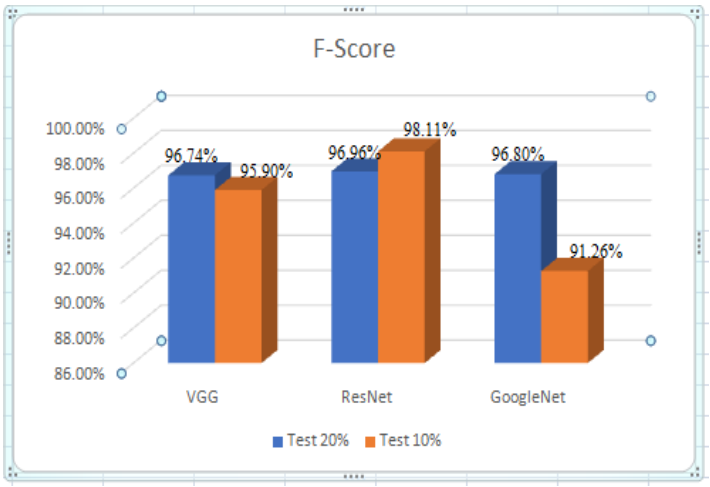


Figure 5: Comparison of F-Score between VGG, ResNet, GoogleNet

4.3. Model Architecture Results

o Intel Core i5-7200U @ 2.50 GHz (4CPUs) Processor

o NVIDIA GeForce 940MX 2010 MB GPU

In Figure 5 the ResNet architecture model with Train-Test 90:10 data has the highest F-Score, which is 98.11%. The VGG architectural model achieved an F-Score of 96.74%, while GoogleNet with Train-Test data of 90:10 had the lowest F-Score of 91.26%.

Figure 6 shown the comparison results of Accuracy between three models, experiments with Train-Test 90:10 data always have a higher level of accuracy compared to Train-Test 80:20, with the exception of the GoogleNet architecture. Of all the architectural models that were tested, the ResNet method with a Train-Test 90:10 produced the highest Accuracy compared to other architectural models, which amounted to 95.94%.

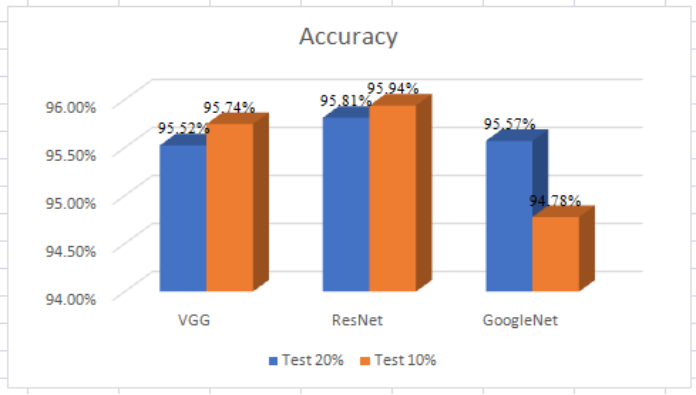


Figure 6: Comparison of Accuracy between VGG, ResNet, and GoogleNet

Two existing approaches are compared to assess the recognition effect of the proposed approach, namely the *Deep CNN – LSTM with Combined Kernels* based approach, and the *NGramCNN* based approach. Table 2 shown the comparison between existing approaches and proposed approaches. From them, it can be seen that the proposed approach achieves has better Accuracy and F-Score, compared to other methods.

Table 2 – Comparison of Accuracy and F-Score by different methods

Method	Accuracy	F-Score
Decision Tree	84.37%	85.41%
Random Forest	86.88%	86.71%
Deep CNN	87.14%	86.45%
NGramCNN	88.73%	89.58%
VGG	95.74%	96.74%
ResNet	95.94%	**98.11%**
GoogleNet	95.57%	96.80%

5. Conclusions

1. The selection of data using different Training and Test Percentages only has a small effect on the value of Accuracy and F-Score, where the difference between the highest Accuracy value between the two percentages is 0.32%, and the difference in the highest F-Score value is 1.14%.

2. The research method used between VGG, ResNet, GoogleNet, and LSTM by using the same parameters only has a small effect on the value of Accuracy and F-Score, where the difference in the highest Accuracy value among all methods is 0.19%, and the difference in the value of F- The highest score ¬ 0.89%.

3. Based on comparison with methods from previous studies, the proposed method has increased from the F-Score results. When compared with the Deep CNN method, the F-Score of the proposed method increased by 11.66, whereas when compared with nGramCNN it increased by 8.53.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] J. H. Kietzmann, K. Hermkens, I. P. McCarthy, B. S. Silvestre, Social media? Get serious! Understanding the functional building blocks of social media, Business Horizon, 2011.

[2] P. Kotler, K. L. Keller, Marketing management, Pearson, 2012.

[3] L. Schweitzer, "Planning and social media a case study of public transit and stigma on twitter", Journal of American Planning Association, **80**(3), 218-238. 2014. https://doi.org/10.1080/01944363.2014.980439.

[4] M. Skuza, A. Romanowski, "Sentiment Analysis of Twitter Data within Big Data Distributed Environment for Stock Prediction", in Proceedings of the Federated Conference on Computer Science and Information Systems, Lodz, Poland, 2015. https://doi.org/10.15439/2015F230.

[5] Y. Kim, "Convolutional neural networks for sentence classification.", 2011, arXiv preprint arXiv:1408.5882.

[6] C. N. Dos Santos, M. Gatti, "Deep convolutional neural networks for sentiment analysis of short texts", in Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, Dublin, Ireland, 2014.

[7] A. Severyn, A.Moschitti, "Twitter sentiment analysis with Deep Convolutional Neural Networks", in Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, 2015. https://doi.org/10.1145/2766462.2767830.

[8] A. Yenter, A. Verma, "Deep CNN-LSTM with combined kernels from multiple branches for IMDB Review Sentiment Analysis.", In 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), New York, NY, USA, 2017. https://doi.org/10.1109/UEMCON.2017.8249013.

[9] E. Cano, M.Morisio, "A deep learning architecture for sentiment analysis" in Proceedings of the International Conference on Geoinformatics and Data Analysis, Prague, Czech Republic, 2018. https://doi.org/10.1145/3220228.3220229.

[10] M. Stowe, Undisturbed REST a guide to designing the perfect API, San Fransisco: MuleSoft, 2015.

[11] B.Liu, Sentiment analysis and subjectivity, Handbook of Natural Language Processing, 2010.

[12] T. C. Carr, R. Hayes, "Social media: defining, developing, and divining", Atlantic Journal of Communication, **23**(1), 46-65, 2015. https://doi.org/10.1080/15456870.2015.972282.

[13] N. Kalchbrenner, E. Grefenstette, P. Blunsom, "A Convolutional Neural Network for Modelling Sentences.", arXiv preprint arXiv:1404.2188, 2014.

[14] K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", 2014, arXiv:1409.1556.

[15] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. arXiv:1512.03385.

[16] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, A. Rabinovich, "Going Deeper with Convolutions.", 2014, arXiv:1409.4842.

[17] F. K. Chopra, R. Bhatia, "Sentiment Analyzing by Dictionary based Approach.", International Journal of Computer Applications, **152**(5). 32-34, 2016. https://doi.org/10.5120/ijca2016911814.

[18] B. Verma, R. S. Thakur, "Sentiment analysis using lexicon and machine learning-based approaches: A survey.", in Proceedings of international conference on recent advancement on computer and communication, Singapore, 2018. https://doi.org/10.1007/978-981-10-8198-9_46.

[19] R. A. Solovyev, M. Vakhrushev, A. Radionov, V. Aliev, A. A. Shvets, "Deep Learning Approaches for Understanding Simple Speech", 2018, arXiv preprint arXiv:1810.02364.

[20] E. Alpaydin, Introduction to Machine Learning, Massachusetts Institute of Technology Press, 2010.

[21] C. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[22] W. Fu, P. O. Perry, "Estimating the number of clusters using cross-validation", Journal of Computational and Graphical Statistics, 1-20, 2019. https://doi.org/10.1080/10618600.2019.1647846.

[23] M. Yulianto, A. S. Girsang, R. Y. Rumagit, "Business intelligence for social media interaction in the travel industry in Indonesia", Journal of Intelligence Studies in Business, **8**(2), 77-84, 2018.

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 5, 287-293 (2019)
www.astesj.com
Special Issue on Multidisciplinary Sciences and Engineering

ASTES Journal
ISSN: 2415-6698

An Integrated Framework for Pronominal Anaphora Resolution in Malayalam

Ajees Arimbassery Pareed*, Sumam Mary Idicula

Department of Computer Science, CUSAT, 682022, India

ARTICLE INFO

Article history:
Received: 13 July, 2019
Accepted: 31 August, 2019
Online: 18 October, 2019

Keywords:
Malayalam
NLP
Anaphora resolution
Deep level tagger
Multilayer perceptron

ABSTRACT

Anaphora resolution is one of the old problems in Natural Language Processing. It is the process of identifying the antecedent of an anaphoric expression in a natural language text. Most of the NLP applications such as text summarization, question answering, information extraction, machine translation etc. require the successful resolution of anaphors. In this paper, we propose a methodology for the resolution of pronominal anaphors present in Malayalam text document. The proposed methodology is a hybrid architecture employing machine learning and rule-based techniques. In our study, we have used a deep level tagger developed using a machine learning based algorithm. The deep level tagger provides detailed information regarding the number and gender of nouns in a text document. The morphological features of the language are effectively utilized for the computational analysis of Malayalam text. Despite using less amount of linguistic features, our system provides better results which can be utilized for higher level NLP tasks such as question answering, text summarization, machine translation, etc.

1 Introduction

Anaphora is a Greek word that originated in 16th-century [1]. It is comprised of 'ana' and 'phora', where 'ana' means back, and 'phora' means carrying. Anaphors can successfully avoid the repetition of a word or phrase in a text document. They frequently occur in both written and spoken forms of the text. Resolving the correct antecedent of an anaphor from the set of possible antecedents is called anaphora resolution. It is one of the complicated problems in Natural Language Processing that has got considerable attention among NLP researchers. Halliday and Hassan defined anaphora as "the cohesion which points back to some previous item" [2]. According to Hirst, "anaphora is a device for making an abbreviated reference to some entity in the expectation that the receiver of the discourse will be able to dis-abbreviate the reference and, thereby, determine the identity of the entity" [3]. Hence, anaphora resolution is the process of disambiguating the antecedent of a referring expression from the set of entities in a discourse. A simple example illustrating the use of an anaphoric expression in English is given below.

Example: ***John*** went to school. ***He*** forgot to take his umbrella.

Here the word '***He***'is the anaphoric expression and '***John***'is the antecedent. Expressions referring to the actual entities in the preceding text are called anaphoric expressions. While the targets of anaphoric expressions are called antecedents.

It is a dream of humankind to communicate with the computers, since the beginning of the digital era. Communication with computers is only possible through a proper understanding of natural language text. The proper understanding of natural language text is possible only after anaphora resolution. There is a large amount of text data available online in Malayalam. Malayalam Wikipedia itself contains more than 30,000 articles. This warrants us to develop tools that can be used to explore digital information presented in Malayalam and other native languages. Anaphora resolution is one such tool that helps in semantic analysis of natural language text. It is also necessary for areas like text summarization, question answering, information retrieval, machine translation, etc.

Anaphora can be broadly classified into three categories [4]. They are pronominal anaphora, Definite noun phrase anaphora, and one anaphora. Among the various types of anaphora, pronominal anaphora is the most widely used one, which is realized through anaphoric pronouns. Pronouns are short words that can be substituted for a noun or noun phrase. They always refer to an entity that is already mentioned in the discourse. Pronominal anaphors are always stronger than full definite noun phrases [5]. They are again classified into five namely personal pronouns, possessive pronouns, reflexive pronouns, Demonstrative pronouns and relative pronouns. All the above-mentioned pronouns need not be anaphoric. A simple example illustrating the use of an anaphoric expression in Malayalam is given below.

*Department of Computer Science, CUSAT, 9061859697 & ajeesap87@gmail.com

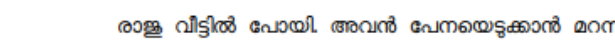


Figure 1: Sample text illustrating the use of an anaphoric expression in Malayalam

Natural languages across the world are quite diverse in their structure. One language does not follow the syntax and semantics of the other. Moreover, anaphora resolution is a challenging task since it requires expertise in different domains of language processing such as lexical analysis, syntactic analysis, semantic analysis, discourse analysis, etc. When it comes to Malayalam, the major challenges associated with anaphora resolution are

- Lack of preprocessing tools
- Lack of standard datasets
- Agglutinative nature of the language
- Free word ordering behaviour
- Influence of cases in resolving the antecedent
- Use of different fonts to encode the text, etc

In this paper, we will focus only on resolution of pronominal anaphors present in a Malayalam text document using deep level tagging algorithms. The structure of this paper is as follows. Section 2 briefly reviews the related works. Section 3 explains the system architecture. Section 4 includes experiments and results. And section 5 concludes the article along with some directions for future work.

2 Related works

Most of the earlier works in anaphora resolution were reported in English and other Europian languages. 'STUDENT' is one of the earliest approaches to resolve anaphors through some automated mechanism. It was a high school algebra problem answering system attempted by Bobrow in 1964 [6]. He used a set of heuristics to resolve anaphors and other questions in a text. The second work in English was reported in 1972 by Winograd [7]. He was the first one to develop an algorithm to resolve pronominal anaphors. In his study, he considered previous noun phrases as the candidate antecedents for the pronominal anaphors. Syntactic positions examine all the possibilities from the preceding text. Wilks introduced a multilevel anaphora resolution system in 1973 [8]. Each level can resolve a set of anaphors based on the complexity to find the antecedent. The first level uses the knowledge of individual lexeme meaning to resolve anaphors. Similarly, the last level uses the "focus of attention" rules to find the topic of the sentence and hence the antecedent.

The modern era on anaphora resolution in English starts with the works of Hobbs in 1977 [9]. He developed an algorithm based on the syntactic constraints of the English language. Parse trees are used to search the antecedent of an anaphor in an optimal order such that the noun phrase upon which the search terminates is regarded as the probable antecedent of the pronoun. Carter reports a shallow processing approach in 1986 [10]. His work exploits the knowledge of syntax, semantics and local focusing. But he didn't give much attention to a large amount of domain knowledge, which he considered as a burden to process accurately. Grosz presented centring theory, a discourse-oriented study for anaphora resolution in 1977 [11]. Lappin and Leass reported another work for anaphora resolution in 1994 [12]. Their algorithm focuses on salience measures derived from the syntactic structure rather than the semantic factors present in the text. McCallum and Wellner presented a CRF based anaphora resolution system for English [13]. Similarly, Charniak and Elsner demonstrated an anaphora resolution system using Expectation Maximisation (EM) algorithm [14].

When it comes to Indian languages, works in anaphora resolution are reported only from few languages such as Hindi, Tamil, Bengali and Malayalam. This situation is contributed by the lack of standard parsers and sophisticated preprocessing tools for Indian languages. The first work in anaphora resolution for Malayalam was reported by Sopha et al. in 2000 as part of her doctoral thesis [15]. Their system (VASISTH) was the first multilingual system for anaphora resolution in Indian languages. Initially, the system was developed for Malayalam and then modified for Hindi. It explores the morphological richness of the language and makes limited use of syntax and other traditional linguistic features. Specific rules are designed for each type of anaphora separately. Athira et al. proposed an algorithm to resolve pronominal anaphors in Malayalam text [16]. They explored a hybrid system for the resolution of pronominal anaphors by incorporating rule-based and statistical approaches. The morphological richness of the language is also utilized in their study. Later, a pronominal anaphora resolution system was proposed by Sobha in 2007 [17]. In her work, she calculated salience weight for each candidate noun phrase from the set of possible candidates. Grammatical features determined the salience factors. The system could achieve comparable performance. High accuracy on a limited data set was the major drawback of her paper. Linguistic issues involved in Co-reference resolution in Malayalam was discussed in detail by Rajendran S in 2018.

3 Proposed method

Our goal is to build an anaphora resolution system for Malayalam which resolves all pronominal anaphors present in a text document. The architecture of the proposed system is shown in figure 2. It shows the general block diagram of pronominal anaphora resolution system for Malayalam. The first module of the architecture is the preprocessing module, where the sentence segmentation and word tokenization operations are carried out. After preprocessing, the preprocessed text is fed to the POS tagging module, where the words from the preprocessed input text are tagged with POS information. The tags used for this study belong to the BIS tagset. The third module of the architecture is the animate noun identification module which identifies the animate nouns from the set of nouns given by the POS tagging module. Animate nouns are the semantic category of nouns referring to persons and animals. They usually correlate with animate pronouns present in a natural language text. Hence identifying animate nouns from the set of nouns in a text document is necessary to resolve the pronominal anaphors present in the document.

After identifying the animate nouns in a POS-tagged text, they

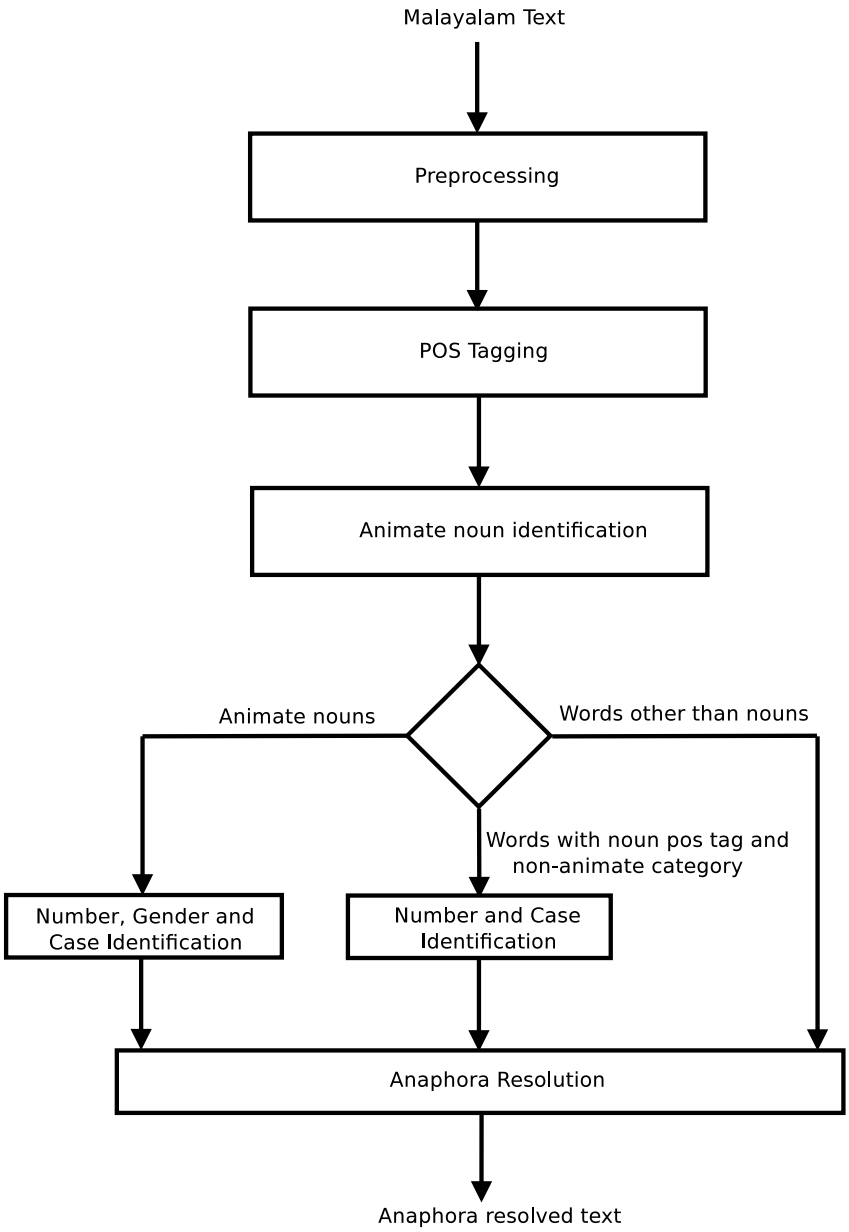


Figure 2: General architecture of the proposed system.

are sent to a number and gender identification neural network. Morphological features of the language are utilized to identify the number and gender information associated with animate nouns. The morphological features are captured with the help of a suffix stripper which can strip the suffixes of different length. Suffixes of different length act as the feature set for number and gender identification module. MLP classifier is used to build the in-depth analyzer of animate nouns. MLP is a class of feedforward artificial neural networks [18]. It is characterized by multiple hidden layers and non-linear activation. The case information associated with the animate nouns is also identified in this module. A rule-based approach is employed for this purpose.

The final module of the architecture is the anaphora resolution module, which is shown in figure 4. The deep level tagged text from the previous modules is the input towards the anaphora resolution module. For each pronoun present in the tagged text, a set of possible candidates are shortlisted. The possible candidates are animate nouns from the previous four sentences of the anaphor, that agree with the number and gender of the pronoun. After shortlisting the possible candidates, a set of salience values are assigned to each candidate. The salience factors are identified through proper analysis of Malayalam text. The set of salience factors we have considered in our study are given in table 2.

The clause information and subject-object information of shortlisted candidates are identified using a rule-based technique. For that, all the sentences containing the shortlisted candidates are divided into clauses. The words with a verb or adjective POS tag act as the clause boundary in a sentence [19]. In the case of adjectives, the nouns following them are also included in the same clause. For each clause identified in the last step, the subject, object and predicate

[['നരേന്ദ്രമോദി\N_NNP', 'സർക്കാരിനെ\N_NN', 'പിടിച്ചുലയ്ക്കുന്ന\V_VM_VNF', 'വിവാദബോംബാണ്\V_VAUX', 'വിജയ്\N_NNP', 'മല്യ\N_NNP', 'ഇന്നലെ\N_NN', 'ലണ്ടനിൽ\N_NNP', 'പൊട്ടിച്ചത്\V_VM_VNF', '.\RDPUNC'], ['ഇന്ത്യയിൽ\N_NNP', 'സാമ്പത്തിക\JJ', 'തട്ടിപ്പുകൾ\N_NN', 'നടത്തിയവർക്ക്\N_NN', 'ബിജെപിയുമായും\N_NN', 'കേന്ദ്ര\JJ', 'സർക്കാരിലെ\N_NN', 'ഉന്നതരുമായും\N_NN', 'അടുത്ത\N_NST', 'ബന്ധമുണ്ടെന്ന\V_VM_VNF', 'ആരോപണം\N_NN', 'പ്രതിപക്ഷ\N_NN', 'കക്ഷികൾ\N_NN', 'നിരന്തരം\RB', 'ഉയർത്തുന്ന\V_VM_VNF', 'ഘട്ടത്തിലാണ്\V_VAUX', 'അവർക്ക്\PR_PRP', 'വെടിക്കോപ്പ\N_NN', 'പകരുന്ന\V_VM_VNF', 'മല്യയുടെ\N_NNP', 'വെളിപ്പെടുത്തൽ-\N_NN', '.\RDPUNC'], ['ഇന്ത്യ\N_NNP', 'വിട്ടം\N_NN', 'മുൻപ്\N_NN', 'ധനമന്ത്രി\N_NN', 'അരുൺ\N_NNP', 'ജയ്റ്റ്ലിയെ\N_NNP', 'കണ്ടിരുന്നു\V_VM_VF', '.\RDPUNC'], ['ചർച്ച\V_VM_VNF', 'നടത്തിയിരുന്നു\V_VM_VF', '.\RDPUNC']

[['നരേന്ദ്രമോദി\N_NNP\ANI', 'സർക്കാരിനെ\N_NN\NANI', 'പിടിച്ചുലയ്ക്കുന്ന\V_VM_VNF', 'വിവാദബോംബാണ്\V_VAUX', 'വിജയ്\N_NNP\ANI', 'മല്യ\N_NNP\ANI', 'ഇന്നലെ\N_NN\NANI', 'ലണ്ടനിൽ\N_NNP\NANI', 'പൊട്ടിച്ചത്\V_VM_VNF', '.\RDPUNC'], ['ഇന്ത്യയിൽ\N_NNP\NANI', 'സാമ്പത്തിക\JJ', 'തട്ടിപ്പുകൾ\N_NN\NANI', 'നടത്തിയവർക്ക്\N_NN\NANI', 'ബിജെപിയുമായും\N_NN\NANI', 'കേന്ദ്ര\JJ', 'സർക്കാരിലെ\N_NN\NANI', 'ഉന്നതരുമായും\N_NN\NANI', 'അടുത്ത\N_NST', 'ബന്ധമുണ്ടെന്ന\V_VM_VNF', 'ആരോപണം\N_NN\NANI', 'പ്രതിപക്ഷ\N_NN\NANI', 'കക്ഷികൾ\N_NN\NANI', 'നിരന്തരം\RB', 'ഉയർത്തുന്ന\V_VM_VNF', 'ഘട്ടത്തിലാണ്\V_VAUX', 'അവർക്ക്\PR_PRP', 'വെടിക്കോപ്പ\N_NN\NANI', 'പകരുന്ന\V_VM_VNF', 'മല്യയുടെ\N_NNP\ANI', 'വെളിപ്പെടുത്തൽ-\N_NN\NANI', '.\RDPUNC'], ['ഇന്ത്യ\N_NNP\NANI', 'വിട്ടം\N_NN\NANI', 'മുൻപ്\N_NN\NANI', 'ധനമന്ത്രി\N_NN\NANI', 'അരുൺ\N_NNP\ANI', 'ജയ്റ്റ്ലിയെ\N_NNP\ANI', 'കണ്ടിരുന്നു\V_VM_VF', '.\RDPUNC'], ['ചർച്ച\V_VM_VNF', 'നടത്തിയിരുന്നു\V_VM_VF', '.\RDPUNC']]

Figure 3: Sample text showing the input and output of the Animate Noun Identification module

Table 1: Distinct classes considered for deep level tagging

Label	Class
M/S	Singular male
M/P	Plural male
F/S	Singular female
F/P	Plural female

information are extracted using the following set of rules [20].

- Predicate —The verb or adjective present in the clause will form the predicate.
- Object —If the predicate is a verb, the noun preceding it will be the object and if the predicate is an adjective, the noun following it will be the object.
- Subject —If the predicate is a verb, the noun preceding the object will be the subject. And if there is no noun preceding the object in the same clause, follow the given rules.
    1. If the predicate of the preceding clause is an adjective, its object will form the subject.
    2. Else the subject of the preceding clause will be the subject.

  Conversely, if the predicate is an adjective, the noun preceding it will be the subject.

Finally, the possible candidates are sorted based on the total salience score and the candidate having the highest salience score is selected as the antecedent of the anaphor.

4 Experiments and Results

In this section, we will discuss the experiments on each phase of the architecture one by one. The first phase is the preprocessing of raw Malayalam text. We have used NLTK implementation sentence tokenizer and word tokenizer for this purpose [21]. Sentence tokenizer converts raw Malayalam text into a list of sentences, and word tokenizer converts each sentence into a list of words. The preprocessed text is provided to the POS tagging module, where the tagged text is generated. A CRF based tagger developed in our department is employed for this purpose [22]. The third module of the architecture deals with animate noun identification. MLP classifier is used to build the animate noun identifier. A set of nouns belonging to both animate and non-animate category is used as the training data. Each noun in the training data is labelled with corresponding class information (animate/non-animate). A set of 109430 nouns were prepared in this way and used for building the machine learning model. 'Fasttext'is used to convert words into vectors. 'Fasttext'is an efficient way to convert words to vectors developed by Facebook's AI research lab [23].

The different configuration of the neural network with various word vector sizes was attempted to build the animate noun identification model. A four-layered architecture with two hidden layers seemed to be the optimum one. The first layer is the input layer which accepts the input. Second and third layers are hidden lay-

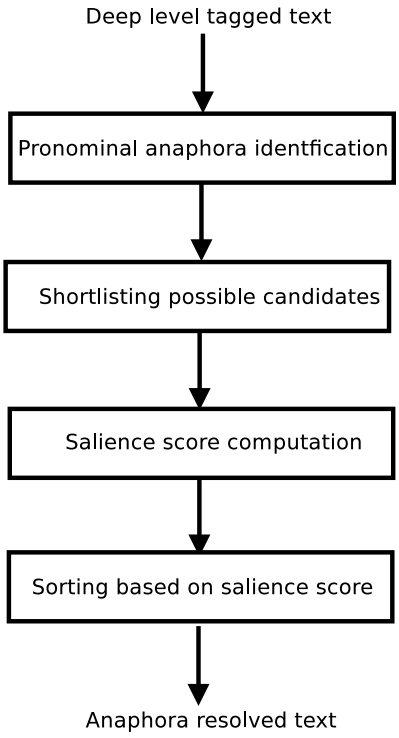


Figure 4: Architecture of anaphora resolution module

Table 2: Salince factors and their weights

Salience factors	Weights
Current sentence	100
Preceding sentences	Reduce sentence score by 10
Subject	80
Object	50
Case	10 to 40
Current clause	100
Immediate clause	50

ers with 'tanh' activation. The last layer is the output layer which generates the output. Softmax is used as the activation function in the output layer. A network with more hidden layers could not improve the performance and only delayed the convergence time. Hence we finalized our hidden layer size as (2,200), where 2 is the number of hidden layers, and 200 is the size of the hidden layers. The maximum accuracy obtained by the classifier on test data is 90.01%. Figure 3 shows the input and output of the animate noun identification module on sample text.

The fourth module of the architecture deals with deep level tagging of animate nouns. MLP classifier is used to build the deep level tagging module. Training data required for the study is a set of names with different number and gender features. Distinct classes considered for the investigation are shown in table 1. 12600 nouns belonging to different categories were prepared for this purpose. Suffixes of various length are used as the feature set for the classifier. Since words and suffixes are symbolic units, it can't be directly fed into neural networks. Hence, we converted our feature set numeric values using Dictvectorizer, a python functionality [24]. Various configurations of the MLP classifier are attempted in our study. A smaller network was not able to represent the data efficiently and increasing the network size did not improve accuracy. Hence, we experimentally finalized our hidden layer size as (2,200), where 2 is the number of hidden layers, and 200 is the size of each hidden layer. 'Relu' is used as the activation function and 'Adam' as the optimizer. The maximum accuracy was obtained by the classifier when the number of features was eight. The performance of the number and gender classifier with a different number of features is shown in figure 5.

In order to evaluate the performance of the proposed algorithm, experiments were conducted on two different types of datasets. These experiments were aimed to find the contribution of animate noun identifier and deep level tagger of nouns to the overall accuracy of the system.

4.1 Dataset1

This dataset contains short stories from children story domain. We have taken short stories in the Malayalam language from [25], a popular site for Malayalam short stories. Each story contains around 20 to 30 sentences of moderate length. Since children stories follow a

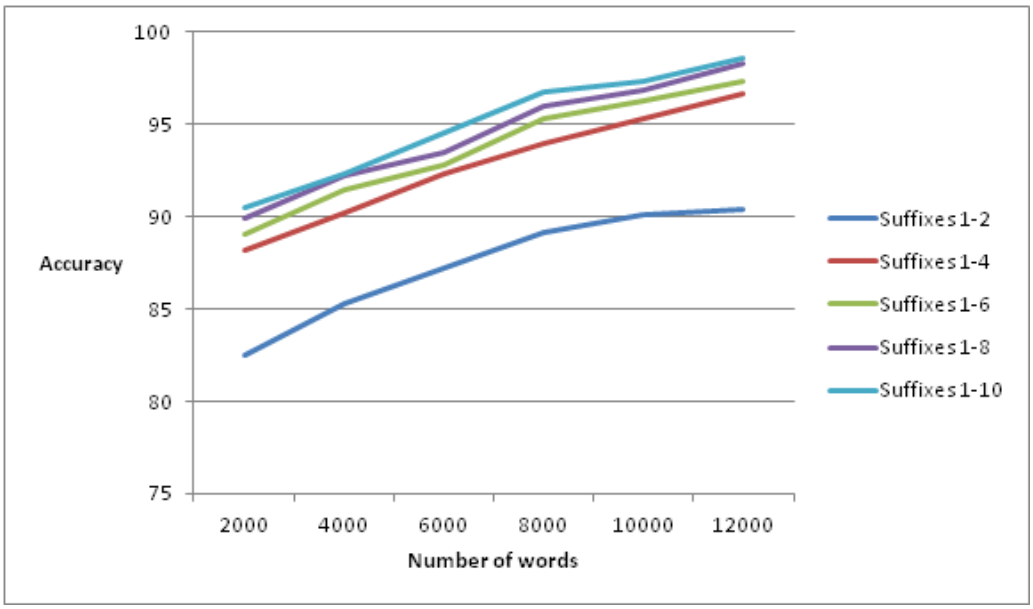


Figure 5: Effect of different features on the performance of the number and gender classifier

Table 3: Results of our experiments on dataset 1

Data set	Total sentences	Total pronouns	Correctly resolved	Accuracy
1	353	19	16	84%
2	466	21	17	80%

straight forward narrative style, the structural complexity of the sentences will be limited. The performance of the anaphora resolution algorithm on this dataset can represent the baseline performance.

4.2 Dataset2

This dataset contains text from news article domain. We have taken news articles from [26], a popular site for Malayalam news. Unlike the children short stories, here the structural complexity of the sentences is not limited. The length of the sentences may go over 25 words. Each article contains around 30 to 40 sentences. Results of our algorithm on both the datasets are summarized in table 3.

From the above experiments, it is observed that the proposed system achieves an overall accuracy of 82.5%. The correctness of the results obtained is verified with the help of language experts. Malayalam is a free word order language, which affects the accuracy of the system. It is observed that the case information associated with the nouns can also decide the antecedent of the pronouns. Some pronouns refer to both animate and inanimate things, which also affects the performance of the system.

5 Conclusion

This paper presents the experimental results of anaphora resolution system in Malayalam language using a hybrid approach. The exclusive feature of the proposed system is the use of machine learning as well as the rule-based technique for the resolution of pronominal anaphors. Malayalam is an agglutinative free word order language, and hence it has many complications in resolving pronouns. The deeper level analysis of nouns helps in the semantic understanding of natural language text. The power of word embedding is also exploited in the study, which helped in recognizing the animate nouns from non-animate nouns. Experimental results show that the deep level analysis of nouns is essential for resolving the pronominal anaphors present in a text document. The morphological richness of Malayalam language is also utilized in this study with the help of suffix stripping algorithms. In our experiments, we have observed that the increase in morphological features results in the increase in accuracy of tagging systems. Hence, incorporating morphological features in the analysis of natural language text appears to be promising for languages such as Malayalam. In the current work, we have focussed only on the resolution of pronominal anaphora. In future, we aim at the resolution of non-pronominal anaphora such as event anaphora, one-anaphora, etc.

References

[1] Kath. Word of the day – anaphora. 2018 (accessed December 7, 2018).

[2] MAK Halliday and R Hasan. 1976: Cohesion in english. london: Longman. 1976.

[3] Grahame Hirst. Anaphora in natural language understanding. 1981.

[4] M Sadanandam and D Chandra Mohan. Telugu pronominal anaphora resolution.

[5] Dan Jurafsky and James H Martin. *Speech and language processing*, volume 3. Pearson London, 2014.

[6] Daniel G Bobrow. A question-answering system for high school algebra word problems. In *Proceedings of the October 27-29, 1964, fall joint computer conference, part I*, pages 591–614. ACM, 1964.

[7] Terry Winograd. Understanding natural language. *Cognitive psychology*, 3(1):1–191, 1972.

[8] Yorick Wilks. Preference semantics. Technical report, STANFORD UNIV CA DEPT OF COMPUTER SCIENCE, 1973.

[9] Jerry R Hobbs. Pronoun resolution. *ACM SIGART Bulletin*, (61):28–28, 1977.

[10] David Maclean Carter. *A shallow processing approach to anaphor resolution*. PhD thesis, University of Cambridge, 1986.

[11] Barbara J Grosz. The representation and use of focus in dialogue understanding. Technical report, SRI INTERNATIONAL MENLO PARK CA MENLO PARK United States, 1977.

[12] Shalom Lappin and Herbert J Leass. An algorithm for pronominal anaphora resolution. *Computational linguistics*, 20(4):535–561, 1994.

[13] Ben Wellner, Andrew McCallum, Fuchun Peng, and Michael Hay. An integrated, conditional model of information extraction and coreference with application to citation matching. In *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, pages 593–601. AUAI Press, 2004.

[14] Eugene Charniak and Micha Elsner. Em works for pronoun anaphora resolution. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 148–156. Association for Computational Linguistics, 2009.

[15] L Sobha and BN Patnaik. Vasisth: An anaphora resolution system for indian languages. In *Proceedings of International Conference on Artificial and Computational Intelligence for Decision, Control and Automation in Engineering and Industrial Applications*, 2000.

[16] S Athira, TS Lekshmi, RR Rajeev, Elizabeth Sherly, and PC Reghuraj. Pronominal anaphora resolution using salience score for malayalam. In *2014 First International Conference on Computational Systems and Communications (ICCSC)*, pages 47–51. IEEE, 2014.

[17] L Sobha. Resolution of pronominals in tamil. In *2007 International Conference on Computing: Theory and Applications (ICCTA'07)*, pages 475–479. IEEE, 2007.

[18] Pedregosa Fabian, Varoquaux Gaël, Gramfort Alexandre, Michel Vincent, Thirion Bertrand, Grisel Olivier, Blondel Mathieu, Prettenhofer Peter, Dubourg Vincent, Vanderplas Jake, et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[19] MS Bindu and Sumam Mary Idicula. A hybrid model for phrase chunking employing artificial immunity system and rule based methods. *International Journal of Artificial Intelligence & Applications*, 2(4):95, 2011.

[20] Rajina Kabeer and Sumam Mary Idicula. Text summarization for malayalam documents—an experience. In *2014 International Conference on Data Science & Engineering (ICDSE)*, pages 145–150. IEEE, 2014.

[21] Edward Loper and Steven Bird. Nltk: The natural language toolkit. In *Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics-Volume 1*, pages 63–70. Association for Computational Linguistics, 2002.

[22] AP Ajees and Sumam Mary Idicula. A pos tagger for malayalam using conditional random fields.

[23] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146, 2017.

[24] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

[25] Bed time stories - wonderful stories in malayalam for kids. `https://kuttykadhakal.blogspot.com/`. Accessed: 2018-09-30.

[26] Madhyamam. `https://www.madhyamam.com/`. Accessed: 2018-08-28.

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 294-299 (2019)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Spectrum Requirements for Control and Non-payload Communication of Maritime Autonomous Surface Ship

Ho Namgung*

Graduate School of Maritime Transportation System, Mokpo National Maritime University, Mokpo, KS005, Republic of Korea

ARTICLE INFO

Article history:
Received: 30 July, 2019
Accepted: 06 September, 2019
Online: 15 October, 2019

Keywords:
MASS
CNPC
LTE-M
ITU-R

ABSTRACT

This paper proposed a frequency bandwidth required for the Control and Non-Payload Communications (CNPC) in autonomous operations of the Maritime Autonomous Surface Ship (MASS). The issue was to seek a radio spectrum to meet future demand for safe operations of autonomous ship at sea. As described in Document 5B/841 of ITU-R SG, safe navigation of autonomous ships relies on communications for command and control. Based on the operations required for the safe navigation of the MASS, the total bit rates were calculated. The spectrum requirements were obtained for command and control communications in operating the MASS safely. As a result, in terms of bandwidth, transmit rate, and transmit distance, state of the art satellite communication system was suitable for a safe operation of the MASS.

1. Introduction

Recently the Maritime Autonomous Surface Ship (MASS) has received much attentions from maritime institutions and industries. The Maritime Safety Committee (MSC) of the International Maritime Organization (IMO) agreed in its 99th session that the Committee provides an output on "Regulatory scoping exercise for the use of MASS" with a target completion year of 2020 [1]. In 2015, ITU-R Study Groups Working Party 5B has requested member states to study and provide a basis for considering regulatory, technical and operational conditions, for the Control and Non-Payload Communications (CNPC) link between satellite services, fixed service, mobile service, and the Unmanned Maritime Systems (UMS) in maritime environments [2]. The society of maritime industries on behalf of United Kingdom has released the MASS Code of Practice Version 1.0 which generally describes all the operations of the MASS including communications systems. As described in the 95th session of MSC [3], significant academic research and commercial research and development (R&D) have undertaken on technological aspect of the autonomous navigation and collision avoidance systems. In particular, the Maritime Unmanned Navigation through Intelligence Network (MUNIN) project proposed internal & external communication systems based on state-of-the-art satellite communication systems and communication requirements [4],[5]. In Finland, data requirements for wireless transmission regarding the environmental sensors and remote maintenance as well as remote control needs were identified for connectivity challenges of autonomous ships in different environments, including ports, deep sea, and Arctic regions [6].

However, there are issues as follows: (i) did not apply a concrete methodology for securing communication channels which are dedicated for safe operation of the MASS via the CNPC; and (ii) only used state-of-the-art satellite communication system based on communication requirements without calculation of spectrum requirements.

Therefore, in this paper, the spectrum requirements among the available maritime frequency band was calculated to secure appropriate communication channel for safe operation of the MASS by analyzing and utilizing the density of ship traffic. Section 2 described methodology for spectrum requirements, and Section 3 calculated the spectrum requirements. Subsequently, applying the calculated spectrum requirements to the available maritime frequency band, it was validated and discussed in section 4. Finally, a summary of the work and the conclusion drew in section 5.

2. Materials and Methods

2.1. Communication Interfaces for CNPC

The possible communication interfaces for the MASS [7] are shown in Figure 1. The communication interfaces is based on analyses provided through ITU-R WP5B document 5B/841 [2]. The MASS Control Station (MCS) is located at either shore-side

*Corresponding Author: Ho Namgung, Email: ngh2009@mmu.ac.kr

or may be mother ship which controls the MASS for some case. The MCS transmits signals for control and commands, while the MASS sends the MCS her state and also information acquired from onboard sensors - so-called telemetry information. The Vessel Traffic Services (VTS) also communicates with both the MCS and the MASS. As shown in the Figure 1, the navigation by ship can start from voyage planning [8]. At the stage of execution and monitoring, the ship's movement is done in the process of situation awareness through detection and/or identification and sequentially collision avoidance as the result of decision-making by mariner. From viewpoint of safe navigation, the Figure 1 represents the communication interfaces for the safe operations of the MASS. Regardless of the conceptual configuration for the communication interfaces, it is not easy to identify all the information, which is required for a safe navigation. International regulations, The International Regulations for Preventing Collisions at Sea (CORLEGs) [9], The International Convention on Standards of Training, Certification and Watchkeeping for Seafarers (STCW) [10], and The International Convention for the Safety of Life at Sea (SOLAS) [11], can be referred to extract the information required for safe navigation by mariner. However, such regulations describe qualitatively the rules for the collision avoidance among ships and the prevention of any maritime accidents. As presented in Table 1, the international regulations illustrate the articles to be referred to identify the information for a safe navigation. Experiences from mariners can represent the information in detail.

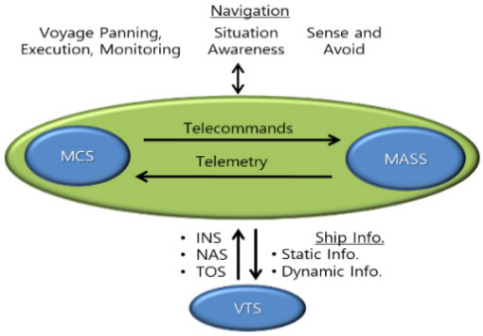


Figure 1: Communication interfaces for safe operations of MASS [7]

Table 1: Regulation related to information for safe navigation

Regulations	CORLEGs	STCW	SOLAS
Articles	Part B. Steering and sailing rules: Rules 5~19	Code A	Chapter V. Safety of Navigation

2.2. Bandwidth for CNPC

To determine the bandwidth for the MASS at sea, a proper frequency band should be identified among available maritime frequency bands. In particular, the CNPC link requires internationally agreed frequency band. This preliminary study provides useful information to find out a proper frequency for the MASS. As described in an Industry Code of Practice, the MASS operations will largely depend on communications systems for control and monitoring of the vessel [12]. It is natural that the CNPC link requires stable and seamless communication. The level of the MASS control will influence on the requirements for the CNPC link. In considerations of the importance of the CNPC link, the two frequency bands, i.e., the Very High frequency (VHF) band, the Ultra High frequency (UHF) band, and the satellite band were being discussed. The present study thus considered the Long Term Evolution - Maritime (LTE-M) for calculating the spectrum requirements of the CNPC.

- LTE-M

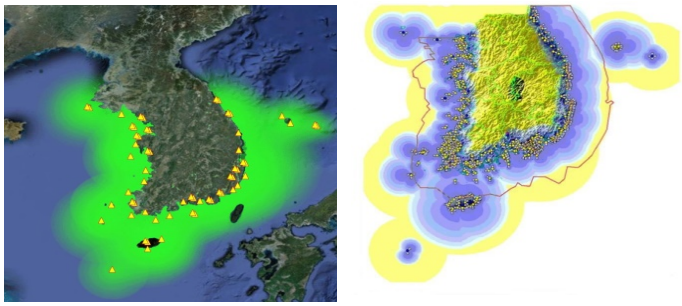
In South Korea, in order to ensure maritime information sharing and communication via access to internet at sea, SMART-navigation project is conducting [13],[14], and [15]. It aims to establish high speed wireless maritime network that covers 100km from offshore through ocean environment conscious cell planning. To facilitate a seamless data transfer various services, therefore, LTE-M is constructing. It is 700 MHz of the UHF band, around the harbor for a variety of services, depending on the strategy to a minimum 1 Mbps transfer rate communication network with the appropriate volume up to 100 km. It enables the free exchange of multimedia data between ships and ship and shore and free use of internet in port and coastal area. In addition, the frequency band of The Downlink (DL) and the Uplink (UL) have10 MHz, respectively.

2.3. AIS

Automatic Identification System (AIS) [16] is an automatic tracking system that uses transponders on ships and is used by VTS. It can transmit information between ships and between ships and land station within range of 50 nautical miles (NM) via VHF in the 160 MHz band. Ships fitted with AIS transceivers can be tracked by AIS base stations located along coast line. Information contains the ship's position, course, speed, name, type, destination, and dimension. SOLAS requires AIS to be fitted aboard international voyaging ships with 300 or more gross tonnage (GT), and all passenger ships regardless of size. A total of 42 AIS base station were constructed on the coasts of South Korea. In present study, in order to take into consideration analysis of the ship traffic flow, data was obtained via AIS during period from January 1 to February 24, 2016.

2.4. Coverage cell

As mentioned in section 2.2, LTE-M enables ships to communicate and exchange data within 100 km (i.e., about 50 nm) of coastal waters from the shore. However, LTE-M with high speed rates for maritime communication network is presently constructing. Therefore, density of the MASS per coverage cell was estimated based on the AIS base stations with about 50 nm of detection distance by considering communication distance and domains as shown in Figure 2.



(a) AIS coverage (b) LTE-M coverage

Figure 2: Comparison of AIS coverage and LTE-M coverage [13],[14], and [15]

2.5. Methodology for Spectrum requirements of CNPC

It was utilized based on '*Characteristics of unmanned aircraft system and spectrum requirements to support their safe operation in non-segregated airspace*' [17] due to non-existence of methodology for spectrum requirements of the MASS's CNPC. In order to calculate spectrum requirements of the CPNC, four step has been composed as presented in Figure 3: (i) analyzed the information exchange rates per the MASS based on the CNPC; (ii) estimated the density of the MASS in coverage cell (e.g., AIS base stations area) per hour; (iii) calculated the spectrum requirements for each type of coverage cell based on (ii) and the frequency efficiency by LTE-M; and (iv) multiplied the reuse factor for each cell type, then determined a total of the spectrum requirements for the CNPC.

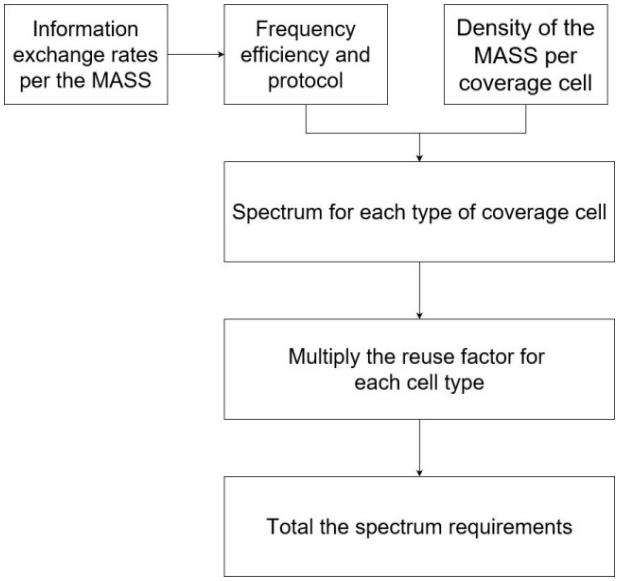


Figure 3: Methodology for spectrum requirements of the CNPC

Calculation methodology of the spectrum requirements for the CNPC of the MASS is shown in equation (1).

$$W_{mass} = K \times B \times M \times R / U \times E \quad (1)$$

Where W_{mass} denotes spectrum requirements, K is reuse factor, B is data rate per the MASS, M is the number of the MASS per cell, R is redundancy factor, U is utilization factor, E is frequency efficiency.

3. Calculation factor for spectrum requirements of CNPC

3.1. Data rate for CNPC (B)

Based on the communication interfaces, a basic data rates is provided shown in Table 2. To calculate the basic data rates, the information required to ensure a safe navigation has been extracted by taking the international regulations into account as mentioned in section 2.1. The DL data is transmitted from MCS or VTS to the MASS, while the UL data is delivered to the MCS or the VTS from the MASS. From Table 2, each data channel needs at least 769.4 Kbps for the UL, and 78.3 Kbps for the DL.

Table 2: Basic data rates for a safe navigation

Telecommand & Telemetry		VTS Communications				Sensing & Collision Avoidance				
		Voice Relay		Data		AIS		RADAR	Target Tracking	Image, Video
DL	UL	DL	UL	DL	UL	DL	UL	UL	UL	UL
0.1742	0.262	64	64	4.53	0.0789	9.6	4.8	100	0.296	600
Total DL(Kbps)		78.3								
Total UL(Kbps)		769.4								

3.2. Density of the MASS per coverage cell (M)

According to the development trend all over the world, the MASS aimed to launch short autonomous runs in 2020 and completely unmanned ships on the high seas by 2025. However, no one knows how many of the MASS would be navigated at sea. Therefore, this study supposes that the MASS would be operated on behalf of present navigating ships at sea. At this time, considering the traffic flow of navigating ships in sea area of South Korea, the number of the MASS was determined.

Among sea areas of South Korea, the Janganseo sea area including the Incheon sea area (the Incheon port, the Pyeongtaek port, and the Daesan port) was considered for analyzing the density of ship traffic. In particular, Janganseo sea area has a lots of ship traffic due to pilot station, anchorage, and junction of inbound and outbound ships. Figure 4 presents the detection range of the Incheon sea area.

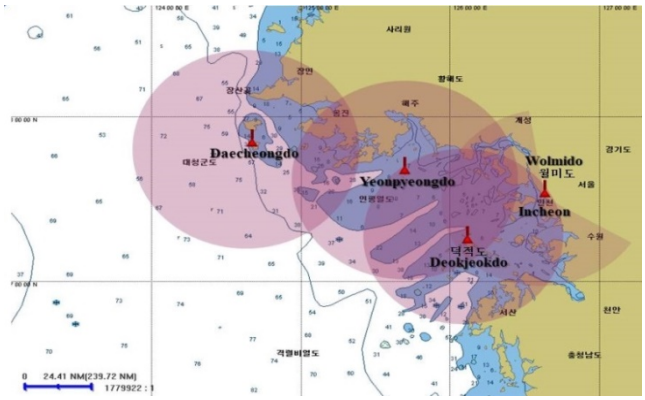


Figure 4: AIS coverage in the Incheon sea area

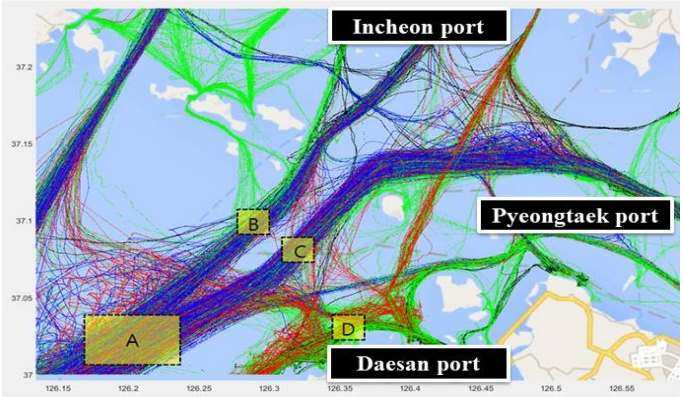
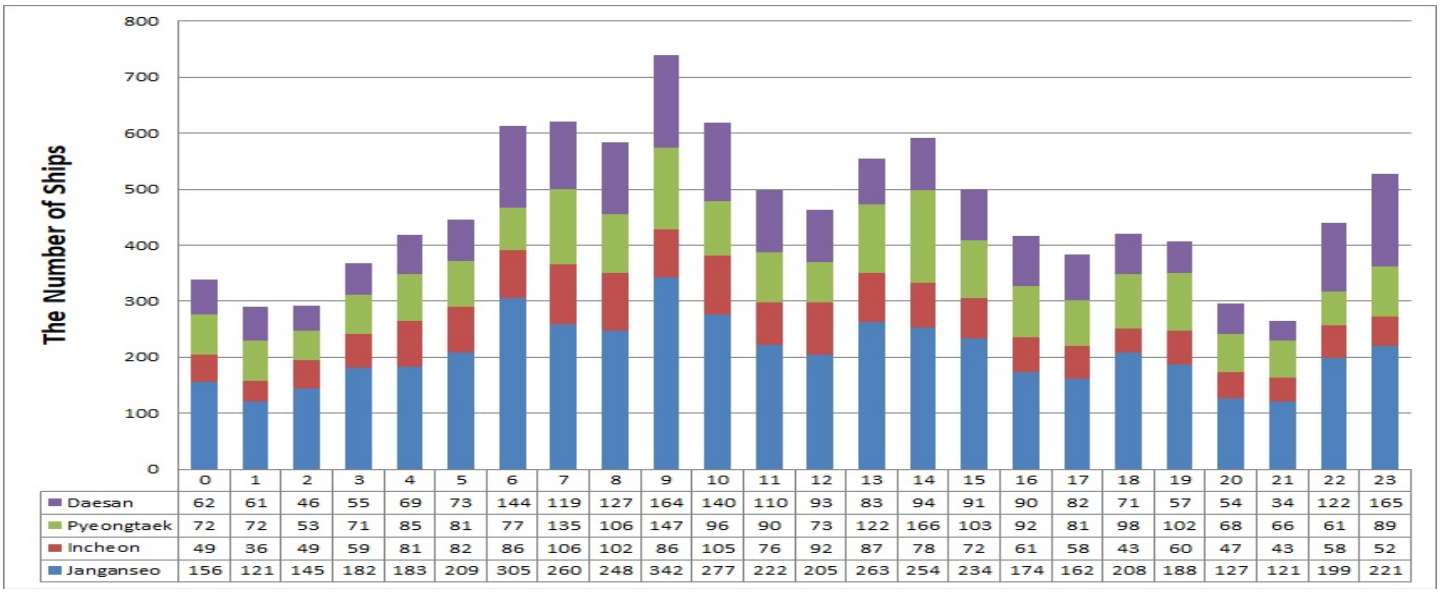


Figure 5: Distribution of ship traffic in the Incheon sea area



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Daesan	62	61	46	55	69	73	144	119	127	164	140	110	93	83	94	91	90	82	71	57	54	34	122	165
Pyeongtaek	72	72	53	71	85	81	77	135	106	147	96	90	73	122	166	103	92	81	98	102	68	66	61	89
Incheon	49	36	49	59	81	82	86	106	102	86	105	76	92	87	78	72	61	58	43	60	47	43	58	52
Janganseo	156	121	145	182	183	209	305	260	248	342	277	222	205	263	254	234	174	162	208	188	127	121	199	221

Figure 6: Distribution of ship traffic per hour in the Incheon sea area

The number of ships was identified by analyzing the ship traffic statistics by time of day based on traffic density. At this time, it was measured the frequency of ship traffic during the most frequent hours to determine the number of ships per cell. As shown in Figure 5, the Janganseo sea area is distributed as follows: (a) Janganseo sea area; (b) Entrance sea area of Incheon port; (c) Entrance sea area of Pyeongtaek port; and (d) Entrance sea area of Daesan port. Figure 6 shows distribution of ship traffic per hour in the Incheon sea area.

During period from January 1 to February 24, 2016, the highest frequency of ship traffic was at 09:00 hours. The number of ships in the Janganseo sea area (a), the number of entrance sea area of Incheon port (b), the number of entrance sea area of Pyeongtaek port (c), and the number of entrance sea area of Daesan port (d) were 342, 86, 147, and 164, respectively. The Sum of (a), (b), (c), and (d) was 739. And these values were divided by observation period (i.e., from January 1 to February 24) for obtaining the average number of ships in the Incheon sea area. As a result of calculation, 13.4 were measured per hour.

3.3. Redundancy (R) and utilization factor (U)

The Redundancy Factor is information that adds additional information to the original information to ensure accuracy during data transmission. Utilization Factor is the ratio used for each function. Table 3 presents the values of the redundancy factor and utilization factor according to [17].

Table 3: Value of the redundancy and utilization factor

Factor	Telecommand & Telemetry	VTS Communication	Sensing & Collision Avoidance		
			Target tracking	RADAR, AIS	Video, Image
R	2	2	2	1	1
U	0.5	1	0.5	1	0.5

3.4. Frequency efficiency (E)

As the mobile communication network evolves, the frequency efficiency improves. The frequency efficiency refers to the amount of data bits that can be processed for 1 second at a frequency of 1 Hz in a cell. As the frequency efficiency increases, the frequency capacity increases. In order to calculate the required mobile broadband the spectrum requirement within the next 10 years in South Korea, we analyzed the devices that can use mobile communication network within the present or the next 10 years, and estimated frequency efficiency by year [18]. Table 4 shows the annual frequency efficiency according to the LTE technology. According to [18], the present study measured the spectrum requirements by using the frequency efficiency 1.30 in 2020.

Table 4: Prospect of frequency efficiency for LTE

	Prospection				
LTE	2016	2017	2018	2019	2020
	0.97	0.97	1.08	1.19	1.30

3.5. Frequency reuse factor (K)

It is a technology used for the allocated frequency to increase the frequency utilization efficiency. In the case of mobile communication, the service area is divided into a plurality of cells, adjacent cells use different frequencies, and non-adjacent cells use the same frequency to increase frequency utilization efficiency. Table 5 shows the frequency reuse factor according to the shape of the cell.

Table 5: Frequency reuse factor according to the shape of the cell

Cell type	Cell radius (Km)	K applied
A	65	7
B	157	4
C	315	3
D	480	3

We thus applied the cell radius 65 km of cell type 'A' due to taking into consideration AIS coverage and LTE-M coverage as presented in section 2.4.

Table 6: Spectrum requirements for CNPC of the MASS

Factor	Telecommand & Telemetry		VTS communications				Sensing & Collision Avoidance				
			Voice Relay		Data		AIS		RADAR	Target Tracking	Image, Video
	DL	UL	DL	UL	DL	UL	DL	UL	UL	UL	UL
K	7	7	7	7	7	7	7	7	7	7	7
B (bps)	174.2	262	64,000	64,000	4,530	78.9	9,600	4,800	100,000	296	600,000
M	13.4	13.4	13.4	13.4	13.4	13.4	13.4	13.4	13.4	13.4	13.4
R	2	2	2	2	2	2	2	2	1	2	1
U	0.5	0.5	1	1	1	1	1	1	1	0.5	0.5
E	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3
Hz	50,276.8	75,617.2	9,235,692.3	9,235,692.3	653,713.8	11,385.8	1,385,353.8	692,676.9	7,215,384.6	85,430.1	86,584,615.3
Total DL			11.3 MHz								
Total UL			103.9 MHz								

4. Results and Discussion

4.1. Results

Spectrum requirements of the CNPC was measured according to equation (1). As shown in Table 6, total DL, which is transmitted from MCS or VTS to the MASS, was about 11.3 MHz, and total UL, which is delivered to MCS or the VTS from the MASS was about 103.9 MHz. Accordingly, a total of the spectrum requirements of the CNPC was about 115.2 MHz. At this time, the number of the MASS navigated in further was considered as 13.4 according to section 3.2. Therefore, total DL and total UL would be spent on about 0.8 MHz and about 7.8 MHz, respectively, in case of using a single MASS with the transmit rate 0.8 Mbps.

In terms of the three category: (i) Telecommand and Telemetry; (ii) VTS communications; and (iii) Sensing & Collision avoidance, the amount of the utilization rate out of the total spectrum requirements was observed as presented in Tables 7 and 8. In the utilization rate of the DL, it was indicated that the utilization rate of the VTS communications was higher than other category. On the other hand, in case of the UL, Image and Video were observed as the highest utilization rate. In particular, the total of DL and UL measured from Telecommand and Telemetry was the smallest when comparing of other category.

Table 7: Utilization rate of DL

Telecommand & Telemetry		DL	0.44 %
VTS communications	Voice Relay	DL	81.55 %
	Data	DL	5.77 %
Sensing & Collision Avoidance	AIS	DL	12.24 %
Total DL			100 %

Table 8: Utilization rate of UL

Telecommand & Telemetry		UL	0.07 %
VTS communications	Voice Relay	UL	8.88 %
	Data	UL	0.03 %
Sensing & Collision Avoidance	AIS	UL	0.68 %
	RADAR	UL	6.96 %
	Target Tracking	UL	0.08 %
	Image, Video	UL	83.3 %
Total UL			100 %

4.2. Discussion

In section 4.1, it could be seen that the spectrum requirements of the CNPC and the utilization rate of the UL and the DL were able to be calculated according to [17]. In this section, securing communication channels for a safe operation of the MASS in terrestrial communication system (i.e., VHF band, UHF band) and satellite communication system via the result of the spectrum requirements of the CNPC were discussed, as well as some general reviews of the calculation method.

VHF band (i.e., marine frequency 156 to 174 MHz) is most suitable for short distance communication with the MASS or mutual communication between the MASS and manned ship, but transmission speed is within about 50 Kbps per channel. In case of applying to the calculated spectrum requirements to the VHF band, obtained transmit rate 0.8 Mbps may be impossible to operate in VHF band, and bandwidth, for operating the number of the MASS presented in section 4.1, was not sufficient.

As described in section 2.2, the LTE-M used 700 MHz of the UHF band has transfer rate as a minimum 1 Mbps in bandwidth of the UL and DL, 10MHz , respectively. Applying the calculated spectrum requirements of the CNPC to the LTE-M, the number of the MASS, as mentioned in section 4.1, would not be operated at the same time due to excess of the bandwidth. Whereas, a single MASS with the transmit rate 0.8 Mbps can be operated within bandwidth of the LTE-M due to less than 10 MHz of the DL and the UL, respectively.

According to MUNIN project carried out from 2012 to 2015, they noted that the ship control systems should be designed to operate with different levels of communication quality. Three levels are suggested with respectively 4 Mbps, 128 Kbps and low bandwidth rendezvous services via state of the art satellite communication [4],[5]. In addition, Global Xpress (5 to 50 Mbit/s) was considered for available satellite communication link of the MASS. State of the art satellite communication is possible to obtain required bandwidth according to the user's demand. Therefore, as a result of applying the spectrum requirement of the CNPC to the satellite communications, it may be possible to operate the single MASS or the several MASS. Consequently, the satellite communication link may be more efficient than the terrestrial communication link because a number of the MASS can be operated at sea under control of the MCS due to securing the spectrum requirement of the CNPC. However, satellites normally have a more restricted link budget than terrestrial communication system as power must be generated from limited area solar panels and also because transmission distances are much longer.

A drawback with the calculation of the spectrum requirements of the CNPC is that it only relied on [17] because calculation methods for the MASS has been not existed at present. For this reason, in order to obtain the accurate spectrum requirements of the CNPC, it is necessary to measure suitable factors for the MASS (e.g., reuse factor, redundancy factor, utilization factor, and frequency efficiency).

5. Conclusion

In this paper, appropriate maritime frequency bands were analyzed and verified amongst communication systems via the calculation result of the spectrum requirements of the CNPC. In order to obtain the spectrum requirements, following steps was carried out: (i) obtained the information transmit rates of the CNPC per the MASS ; (ii) estimated the density of the MASS per AIS cell in the Incheon sea area; (iii) calculated the spectrum requirements using factors (e.g., reuse factor, redundancy factor, utilization factor, and frequency efficiency); and (iv) determined a total of the spectrum requirements of the CNPC. The results of applying to the existing and state of the art communication system represented that satellite communication system is suitable for safe operation of the MASS, in terms of bandwidth, transmit rate, and transmit distance. However, in case of using the satellite communication system, it is imperative to invest huge budget for operation of the MASS.

A further study is required to obtain the appropriate values of the factors based on characteristics of the MASS operation system in order to measure the accurate spectrum requirements of the CNPC, and it is necessary to demonstrate the calculation result in comparison with experiment via communication systems (e.g., VHF band , UHF band, and state of the art satellite) at real sea area.

Conflict of Interest

The author declares no conflict of interest.

References

[1] IMO MSC., "Regulatory scoping exercise for the use of Maritime Autonomous Surface Ships (MASS)," MSC 99th session, MSC.99/INF.22, London, 2018.

[2] ITU-R SG, "Characteristics of unmanned maritime systems," The Meeting of working party 5B, Document 5B/841, Bucharest, Romania, 2015.

[3] IMO MSC., "The IMO regulatory framework and its application to Marine Autonomous Systems," MSC 95th session, MSC.95/INF.20, London, 2015.

[4] Ørnulf Jan Rødseth, Beate Kvamstad, and Thomas Porathe, "Communication Architeture for an Unmmaned Merchant Ship," IEEE OCEANS 2013, Bergen, Norway, June, 2013.

[5] Ørnulf Jan Rødseth, Åsmund Tjora, and Pål Baltzersen, MUNIN Deliverable D4.5: Architecture specification, available at http://www.unmanned-ship.org, MRTK, MUNIN-FP7 GA-No.314286, Aug. 2013.

[6] Marko Höyhtyä, Jyrki Huusko, Markku Kiviranta, Kenneth Solberg and Juha Rokka, "Connectivity for Autonomous Ships: Architecture, Use Caes, and Research Challenges," 8th International Conference on ICT Convergence, Oct, 2017.

[7] Ho Namgung, "A Study on Concept Design of Communication System for Safe Navigation of Unmmaned Surface Vehicle," Master dissertation. Mokpo National Maritime University. Mokpo, 2016.

[8] Nathaniel Bowditch, *The American Practical Navigator*, 2002 Bicentennial Ed. Pub. No.9, National Imagery and Mapping Agency, Bethesda, Maryland, 799, 2002.

[9] IMO, *The Convention on the International Regulations for Preventing Collisions at Sea, 1972,* 2010.

[10] IMO, *The International Convention on Standards of Training Certification and Watchkeeping for Seafarers, 1978*, *as amended*, Consolidated Ed. 2011.

[11] IMO, *The International Convention for the Safety of Life at Sea*, Consolidated Ed. 2012.

[12] Society of Maritime Industries UK, *An Industry Code of Practice*, A Voluntary Code Version 1.0, 2017.

[13] Ministry of Oceans and Fisheries, *Report of planning Research for the Development of Korea e-navigation*, 2013.

[14] Ministry of Oceans and Fisheries, *National Strategy for the Development and Implementation of e-navigation*, 2013.

[15] Korea Institute of S&T Evaluation and Planning, *The Report of the Preliminary Feasibility Study on the SMART-navigation project*, 2015.

[16] ITU-R, "Technical characteristics for an automatic identification system using time division multiple access in the VHF maritime mobile frequency band," Rec. ITU-R M.1371-5, 2014.

[17] ITU-R, "Characteristics of unmanned aircraft systems and spectrum requirements to support their safe operation in non-segregated airspace" Rep.2171-5, FEB. 2009.

[18] Korea Communication Commision, "A study on additional spectrum forecast for mobile broadband," 2011.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 300-305 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Resonator Influence Simulation of Designed Close-Open Standing Wave Thermoacoustic Engine

Sugiyanto[*,1,2], Samsul Kamal[1], Joko Waluyo[1], Adhika Widyaparaga[1]

[1]*Department of Mechanical and Industrial Engineering, Faculty of Engineering, Gadjah Mada University, 55281, Indonesia*

[2]*Department of Mechanical Engineering, Vocational School, Gadjah Mada University, 55281, Indonesia*

ARTICLE INFO

Article history:
Received: 28 May, 2019
Accepted: 15 September, 2019
Online: 15 October, 2019

Keywords:
SWTE
Length of Resonator Tube
Acoustic Power

ABSTRACT

Thermo-acoustic technology is very potential to be applied to convert heat into another energy source. The complete thermoacoustic engine construction requires a resonator connected to a core where there is a heat exchanger. The thermoacoustic core has a function as an acoustic power generator while the resonator functions as a storage of acoustic power which will shift the amplitude or phase of oscillating pressure or volume flow rate. Experimental testing was carried out to obtain the performance of a quarter wave length of standing wave thermoacoustic engine (SWTE). The resonator is made of a 2 inch Sch 40 stainless steel pipe with a length of 390 mm. The results of this experiment showed that the measured pressure amplitude was 4800 kPa with a working frequency of 138 HZ and produced an acoustic power of 22.85 watts. The parameters of this experimental test are used for the simulation program. The acoustic power generated from the simulation is 20.13 watts, 18.67 watts and 17.82 watts respectively for the length of the resonator 390 mm, 780 mm and 1170 mm.

1. Introduction

Thermoacoustic is a technology related to heat transfer to pressure oscillation, so the use of the term "Thermo" comes from the heat transfer process while "acoustic" comes from pressure oscillating in the working fluid. This term was first introduced by [1] who was credited with laying a theoretical foundation for the field of thermoacoustic research. Thermoacoustic technology is more interesting to develop because of its superiority compared to traditional energy conversion system technologies [2]. The advantage lies in thermoacoustic devices that have immovable components and simple construction (basically pipelines), making such systems very reliable and requires low construction and maintenance costs.

Thermoacoustic devices classified into thermoacoustic engine (TAE) and thermoacoustic refrigerators (TAR) depending on the practical application of the thermoacoustic effects. TAE turns thermal energy into acoustic energy. The thermal energy comes from a significant temperature difference derived from a heat exchanger that applied to two ends of porous material. TAR converts acoustic energy into temperature differences utilizing acoustic waves imposed along with porous material. This process causes temperature differences that occur in the two heat exchangers [3].

The main components of a TAE or TAR are a resonator tube, a porous material (stack) and two heat exchangers. A standing wave type produces a half wavelength or a quarter wavelength which generate in the resonator. The thermoacoustic phenomenon takes place in porous material when the temperature gradient that is worn along the porous material interacts with the oscillation of sound waves. The heat exchanger is responsible for transferring heat in and out of the thermoacoustic device at the desired temperature, thereby maintaining the temperature gradient applied to the porous material [4].

Swift [5] built and tested this SWTE using a 5-inch diameter stainless steel pipe with electric heating. The working fluid used is Helium at an average pressure of 13.8 bar. The result is an acoustic power that can be generated at 630 W with an efficiency of heat conversion to acoustic power of 9% and a Carnot efficiency of 13%. Thermoacoustic prime mover (TAPM) which has a resonator length of 128 cm, uses a stack of stainless-steel mesh numbered 14, two heat exchangers and water at atmospheric pressure as gas acting inside the resonator. Successfully designed, built, and evaluated by [6].

[*]Corresponding Author: Sugiyanto, Department of Mechanical Engineering, Vocational School, Gadjah Mada University, 55281, Yogyakarta, Indonesia. Tel : +628112502021, Email: sugiyanto_t@ugm.ac.id

There are two types of acoustic waves generated by the thermoacoustic engine, standing waves, and traveling waves. Both of these wave types depend on the type of resonator used. Thermoacoustic engines require high-temperature gradients to produce sound. The temperature gradient itself occurs in part often called the stack (for a traveling wave thermoacoustic engine called a regenerator), placed between the hot and cold sides of the thermoacoustic engine. Thermoacoustic engine research mostly is done on the standing wave type due to its ease of configuration, which only consists of a resonator pipe and the thermoacoustic core which is directly connected to the resonator and placed near the other end. This thermoacoustic core is a heat exchanger (HHX), stack, and cold heat exchanger (CHX).

The influence of the length of the resonator pipe on frequency is carried out by [7] who investigate the performance of prime mover thermoacoustic traveling waves at a small onset temperature difference, which is obtained by mounting the stack in the right position on the horizontal resonator pipe.

The resonator tube is one of the critical components of a thermoacoustic engine. Many studies have been conducted on the effect of resonance tube on thermoacoustic performance, most of them focused their research on resonance tube length which is considered the dominant factor that determines the frequency of the system

The effect of the resonance tube on a standing wave thermoacoustic engine is researched by simulating a linear thermoacoustic. In this simulation, the SWTE is connected to a pulse tube refrigeration to obtain the temperature gradient on the core side. The results of this simulation calculation show that the prolongation of the resonance tube can cause a decrease in the working frequency [8].

Hariharan et al. [9] conducted a study with variations in stack geometry and resonator length to determine the SWTE performance. In this study, variations in the length of the resonator carried out with lengths of 200, 300, 400, and 500 mm. Variation of the length of the resonator influences the pressure amplitude, wave frequency, and acoustic power so that it will also affect its efficiency.

The TAE model whose resonator length can be regulated was examined by [10] which consisted of 103 mm ceramic honeycomb stack, volume buffers, and shell-tube type heat exchangers. The TAE performance was evaluated using six different types of stack. The length of the resonators used were 50 mm, 100 mm, and 150 mm placed in three different places. The length of the resonator that can be set in this study aims to choose the best working frequency so that the TAE can work optimally and produce the highest acoustic power.

The influences of the tube diameter of traveling wave thermoacoustic are not sensitive; hence, the resonance frequency of the radial dimension is always ignored. Therefore, the way to adjust the working frequency is to vary the length of the resonator tube. However, for a small thermoacoustic system, the influence of the radial dimension cannot be neglected [11].

The resonator is one of the key components of a thermoacoustic engine because it's save acoustic power. Resonator supply the instances and compliances that the working frequency is related to it. The unwanted eddying or non-linear pressure variations should be avoided by using a straight tubular resonator pipe without step, misalignment, and sudden transition. However, A straight, uniform-diameter resonator is seldom the best design [12].

This paper is an extension of work originally presented in the 2018 4th International Conference on Science and Technology (ICST), by changing some variations but still focusing on simulations as a way to analyze the performance of this SWTE. The author has conducted research on the effects of resonator diameter on SWTE performance by simulations using DeltaEC. The results show that the simulations with resonator diameter variations from the close-open SWTE have no significant effect [13].

Experiment testing was conducted beforehand to get the results of the parameters used. The results of this experiment will be used to validate the results of the simulation. The parameters of the simulation results that have been confirmed with settings in the SWTE experiment are then used as SWTE simulation parameters in variation length of the resonator. Simulation of an SWTE is easy to do using DeltaEC . This program was made by Dr. Bill Ward and Dr. Gregori Swift from Los Alamos National Laboratories (LANL) [14].

2. Theoritical basis

The quarter wave length resonator is a tube closed at one end and open at the other. This resonator can be explained in a lumped element model consisting of compliance in series with inertance [15].

The working frequency is related to the inertance (L) and compliance C) of the component in the thermoacoustic engine, and they can be described by the equation [12]:

$$L = \frac{\rho_m l}{A} \tag{1}$$

$$C = \frac{V}{\gamma P_m} \tag{2}$$

Where A and l are the sectional area and length of the resonator tube, respectively. ρ_m is average density and P_m is the pressure of the gas. V and γ are the volume compliance and the specific heat, respectively.

The acoustic intensity I for the measured sound wave was calculated using the modified two sensor method [16]:

$$I = \frac{1}{8\omega\rho_m}\left\{\mathrm{Im}[H]\left(|P_A|^2 - |P_B|^2\right) + 2\,\mathrm{Re}[H]\,|P_A||P_B|\sin\theta\right\} \tag{3}$$

H is defined by:

$$H = \frac{kF}{\cos\left(k\frac{\Delta x}{2}\right)\sin\left(k\frac{\Delta x}{2}\right)} \tag{4}$$

Where *Re* [ ] is the real component and *Im* [ ] is the imaginary components. P_A and P_B are acoustic pressure waves measured by pressure sensors mounted close together along the resonator, and $\theta = \arg [P_A/P_B]$ represents phase lead of pressure wave P_A relative to P_B, k is complex wave number and Δx is the difference in distance between the two pressure sensors.

F is a complex factor calculated by:

$$F = 1 - \frac{2J_1\left(i^{3/2}\sqrt{2}r_0/\delta\right)}{\left(i^{3/2}\sqrt{2}r_0/\delta\right)J_0\left(i^{3/2}\sqrt{2}r_0/\delta\right)} \quad (5)$$

And

$$k = -ik_0\sqrt{\frac{J_0\left(i^{3/2}\sqrt{2}r_0/\delta\right)}{J_2\left(i^{3/2}\sqrt{2}r_0/\delta\right)}}\sqrt{\gamma + (\gamma - 1)\frac{J_2\left(i^{3/2}\sqrt{2\sigma}r_0/\delta\right)}{J_0\left(i^{3/2}\sqrt{2\sigma}r_0/\delta\right)}} \quad (6)$$

Where γ and σ is notation of the specific heat ratio and Prandtl number, respectively. J_n represent the n^{th} order complex of Bessel function, k_0 is the wave number, and r_0 is the radius of duct.

The results of the acoustic intensity calculation (I) are then used to calculate the acoustic power $(\dot{E})$ in the following equation:

$$\dot{E} = \dot{I}A \quad (7)$$

3. Methodology

This research uses an SWTE that is open at the end of the resonator but closed at the end adjacent to the thermoacoustic core as seen in Figure 1. Resonators are made of 2-inch Sch. 40 stainless steel pipes with a length of 390 mm. Two heat exchanger devices used are made of copper both for hot heat exchangers (HHX) and cold heat exchangers (CHX). Both of these heat exchangers flank the stack made of stainless-steel wire mesh, with a porosity of 88.26%. The heat source for heating the HHX comes from 6 pcs glow plug from an 11 VDC diesel engine, and CHX is cooled using water.

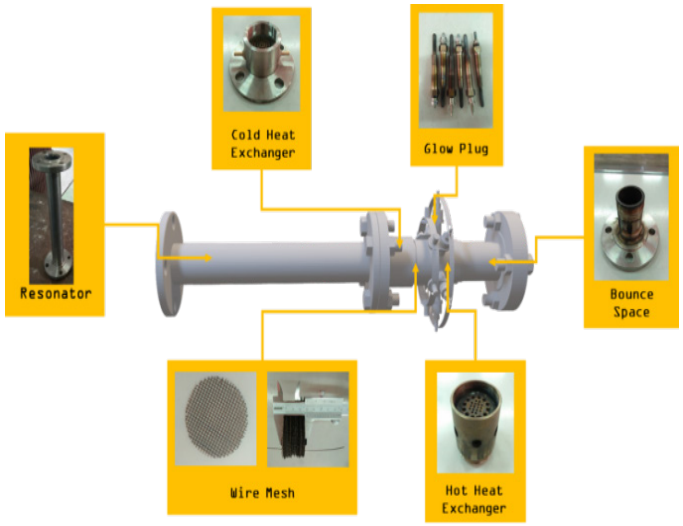


Figure 1: The main components of a close-open thermoacoustic engine

The first step is experimental testing to find out the acoustic intensity of SWTE with a resonator length of 390 mm. In this experimental test, temperatures were measured using two type-K thermocouples placed in HHX (hot side) and CHX (cold side) connect to National Instrument 9234 module. Two M102A05 PCB pressure sensors are mounted on the resonator with a distance between them of 50 mm, used to measure acoustic pressure with readings using the National Instrument module 9234 as well. As a heat source, 6 glow plugs from diesel engines are used which are connected to the transformer battery charger. The set-up of this experimental test is shown in Figure 2.

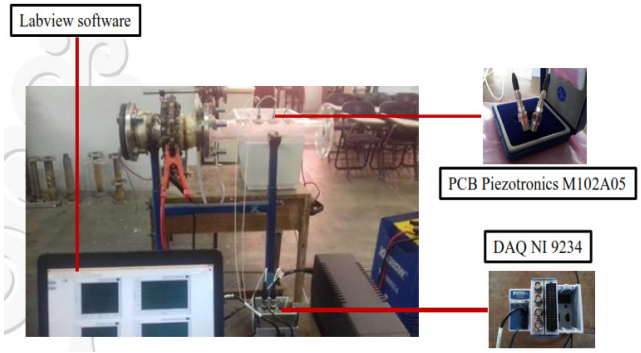


Figure 2: SWTE experiment set-up

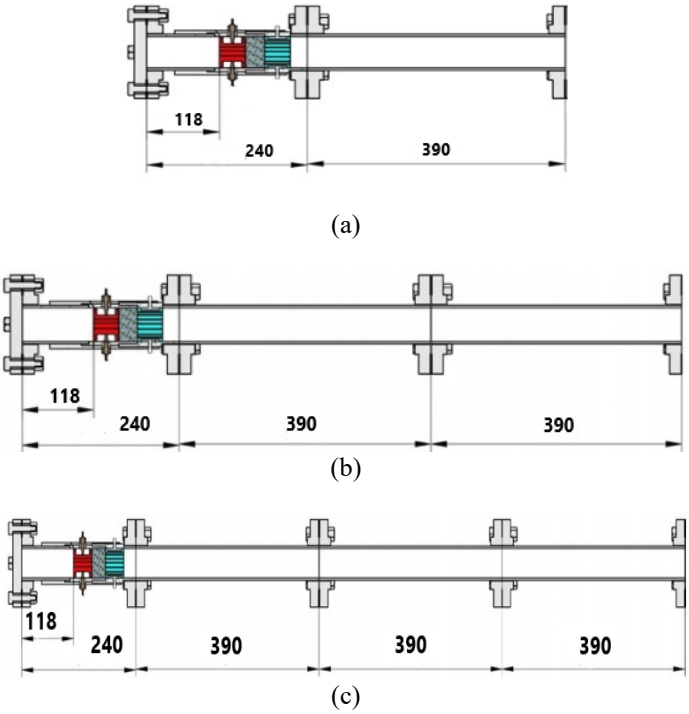


Figure 3: SWTE simulation set-up with a variation of resonator length; a. 390 mm , b. 780 mm , c. 1170 mm length of resonator tube.

The second step is simulation with DeltaEC software to find out the performance of SWTE. The first simulation uses experimental test parameters at a 390 mm resonator length. The experimental data in the first step of the study were used to validate the simulation results in the second step. After the two results between the experiment and the simulation are validated, a simulation program is further developed to predict the performance of the SWTE at variations in the length of 780 mm and 1170 mm as seen in Figure 3.

The display of measurement results using NI 9234 data logger as a simulation parameter that was read by Labview is shown in Figure 4. at point 1 and around 3600 Pa, 138 Hz at point 2. The measurement data from these two pressure sensors are then processed to get the acoustic intensity and acoustic power using equations 3 and 7. The final result for this experiment is the acoustic power of 22,853 watts.

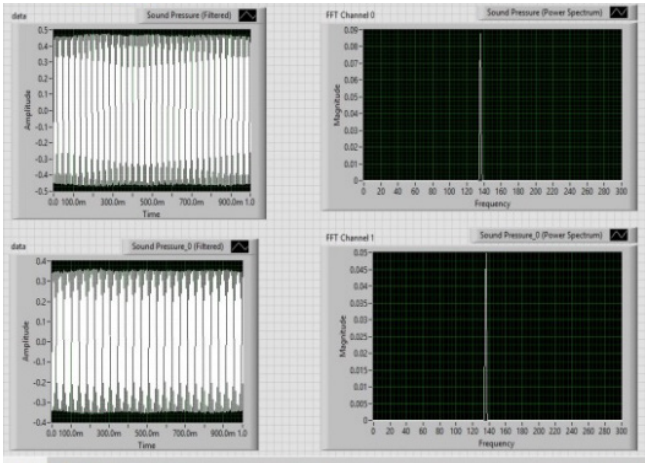


Figure 4: Pressure Amplitude and frequency of SWTE ekperiment result.

4. Result and Analysis

The result of program execution is a schematic SWTE construction diagram with segments of the components used, as presented in Figure 5. It is in a closed on one side state expressed by the SURFACE segment. The other components are Excess pipe, which is stated by DUCT segment, HHX and CHX components are expressed by HX segment, STKSCREEN segments express wire mesh stack, and resonator tube is expressed in DUCT segment.

The DeltaEC program shown in Figure 6 begins by entering the initial parameters as used in the experiment testing, followed by settings for each segment used. In Segment 6 which is the Duct segment for the resonator, a variation of the resonator length of 390 mm is included so that the total power and the acoustic power generated by the SWTE system. The simulation program that has been successfully executed to obtain the performance of SWTE is then used to simulate the performance of SWTE at the length of other resonators. This simulation is done by changing the program simulation parameter at segment DUCT.

The acoustic power that was successfully generated from the simulation for a 390 mm resonator length was 20,13 watts. The difference in acoustic power between experimental results and simulation results is 3%, so this difference is still acceptable because the value of the deviation is even smaller than 10% [12].

The distribution of acoustic power from the system can be seen from the graph in Figure 7 as a result of program execution. The red line shows the total energy on the device while the blue line is the acoustic power. In the segments 1 and 2, which are bounce space areas, there is no power generation. Acoustic power starts to be generated in the core area of SWTE, and the highest is in the field of stack wire mesh which is almost 30 watts. Momentary out of CHX this power goes down and will tend to be constant along the resonator tube which is around 20 watts. The total power of this system is 680 watts, so the efficiency of SWTE with a resonator length of 390 m is 2.94%.

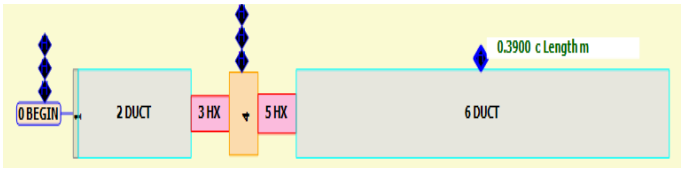


Figure 5: Schematic diagram result of simulation for 390 mm resonator length.

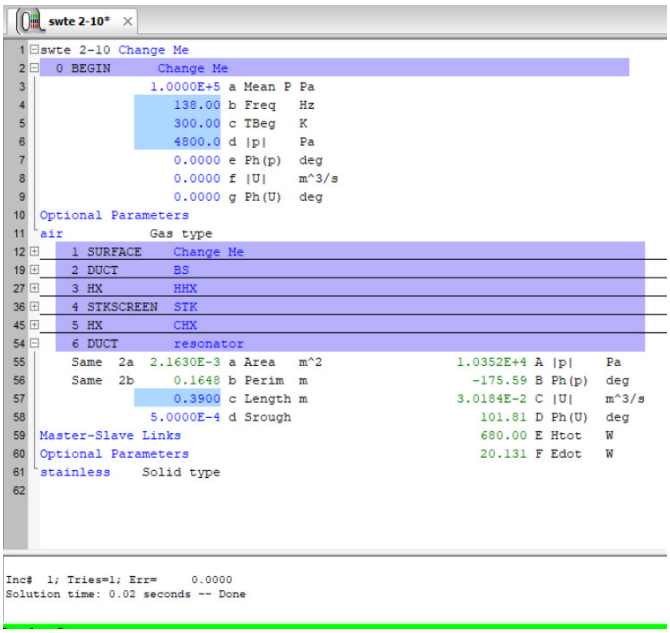


Figure 6: Result simulation of power at 390 mm of resonator length

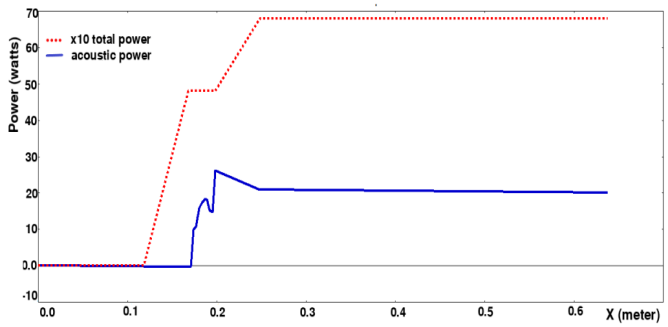


Figure 7: Power distribution of SWTE use 390 mm of resonator length.

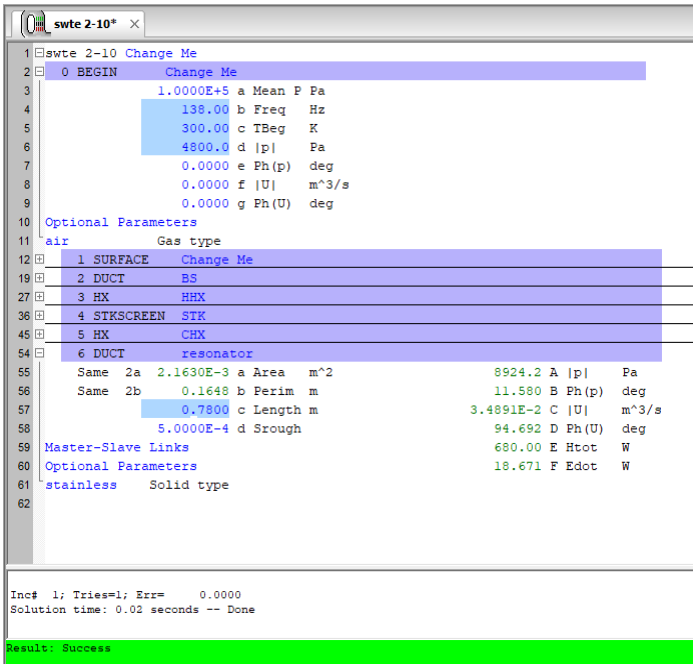


Figure 8: Result simulation of power output at 780 mm of resonator length

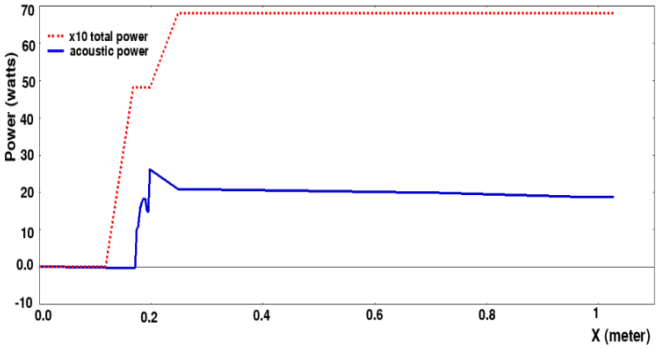


Figure 9: Power distribution of SWTE use 780 mm of resonator length.

Figure 8 to 11 presents acoustic power generated for a resonator length of 780 mm and 1170 mm, respectively. The simulation with a 780 mm resonator length is done by changing the segment 6 in the program by liquefying the other parameters in a fixed condition, as shown in Figure 8. The output power for this condition is 18.67 watts for acoustic power and 680 watts for total power. There was no change in total power due to variations in this second simulation; no other parameters were changed. The efficiency gained from this conditioning system is 2.7%.

As shown in Figure 9, for the length of resonator tube 780 mm at a position of 0 to 0.18 m there is no power generation because a bounce space occupies this position as an air reservoir. At intervals of $0.18 \leqslant x \leqslant 0.2$ m, which is the position of the HHX and stack, the acoustic power starts to rise and reaches the highest value shortly after leaving the stack, which is 26.12 watts. CHX occupies the area at $0.2 \leqslant x \leqslant 0.4$ m intervals, resulting in acoustic power of 20.96 watts. This value is lower than when the stack exits due to water cooling on the CHX side. This value will tend to be constant when passing through the resonator at a distance of $0.4 \leqslant x \leqslant 1.2$ m, which is equal to 18.67 watts.

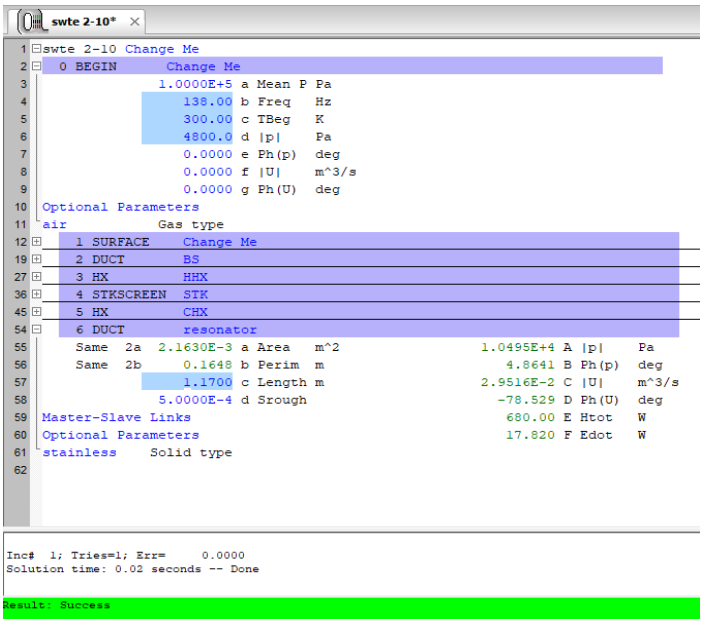


Figure 10: Result simulation of power output at 1170 mm of resonator length

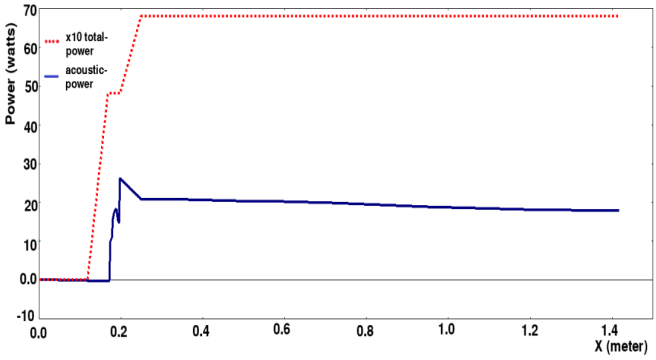


Figure 11: Power distribution of SWTE use 1170 mm of resonator length

The simulation results for the variation of the resonator length of 1170 mm are presented in Figure 10. Like the two previous simulations, segment 6 of the resonator length is changed to obtain information on the amount of output power generated by this system. The results can be seen that the total power produced is still constant at 680 watts because there is no change in the thermoacoustic core parameters. The acoustic power generated is smaller by 17.82 watts so that the efficiency of the system as a result of this third simulation is 2.62%.

The acoustic power distribution as presented in Figure 11 has the same trend with the simulation results for the variation of the resonator lengths of 390 mm and 780 m, where on the bounce space side the end is closed, there is no power generated, the core has the highest value and the resonator will tend to go down until the side open end on this system.

Based on the simulation results for SWTE with variations in the length of the resonator, the results show that the longer the resonator tube, the output power of the system will decrease. The prediction from this simulation is following the research conducted by [9] and [10], namely that the length of the resonator will affect the acoustic power so that it will also affect its efficiency.

The way to increase the standing type of performance can be done by changing to the co-axial type without changing the length of the resonator. Takeyama [17] has tested it by inserting an inner tube into the resonator to the bounce space area. The inner tube used is smaller than the diameter of the resonator tube, as seen in Figure 12. Estimation of co-axial performance can be done by simulations using DeltaEC, such as research by [18] which conducted this simulation to optimize co-axial traveling-wave thermoacoustic cooler powered by the standing-wave thermoacoustic engine.

The simulation results of the co-axial type of thermoacoustic engine which the inner tube diameter of 25 mm and 580 mm length presented in Figure 13. The acoustic power is 30 watts. There was an increase of about 10 watts compared to the SWTE type result in both simulation and experimental. The total energy that can be achieved by co-axial is about 100 watts, so the efficiency of this system is around 3%, slightly higher than the SWTE.

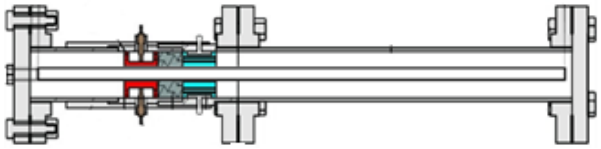


Figure 12: Co-axial thermoacoustic engine contruction.

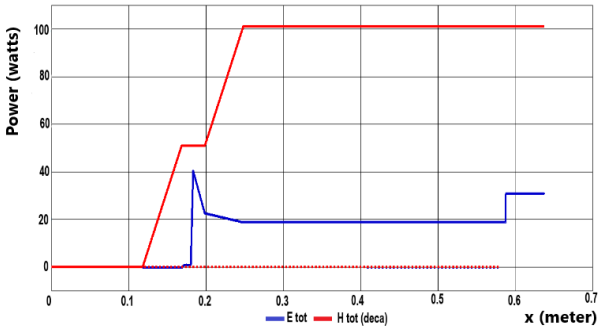


Figure 13: Power distribution of co-axial use 25 mm inner tube diameter

A summary of the comparison of research results in this paper is presented in Table 1

Table 1: Comparison of test results with variations in resonator length

Methods	Length of Resonator (mm)	Acoustic Power (watts)
Experiment	390	22.85
Simulation	390	20.13
Simulation	780	18.67
Simulation	1170	17.82
Simulation (co-axial)	390	30.00

5. Conclusion

Simulations to predict the performance of SWTE open-close have been carried out. Its application for a 390 mm resonator length successfully made, and the results match the experimental data. Execution of other resonator lengths, namely 780 mm and 1170 mm, states that a prolonged resonator reduces the acoustic power produced by this device. Another method for increasing engine performance without changing the length of the resonator is by making the device co-axial type.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The author would like to thank the Gadjah Mada University Higher Education Research Scheme, which has supported part of this research.

References

[1] N. Rott, “Thermoacoustics,” Adv. Appl. Mech., vol. 20, pp. 135–175, 1980. https://doi.org/10.1016/S0065-2156(08)70233-3

[2] S. L. Garrett, “Thermoacoustic Refrigeration,” Naval Postgraduate School, Breckenridge Hilton, Breckenride CO, 1993.

[3] G. . Swift, “Thermoacoustic engines,” J. Acoust. Soc. Am., vol. 84, no. 4, pp. 1145–1180, 1988. https://doi.org/10.1121/1.396617

[4] H. Babaei and K. Siddiqui, “Design and optimization of thermoacoustic devices,” Energy Convers. Manag., vol. 49, no. 12, pp. 3585–3598, 2008. https://doi.org/10.1016/j.enconman.2008.07.002

[5] G. W. Swift, “Analysis and performance of a large thermoacoustic engine,” J. Acoust. Soc. Am., vol. 92, no. 3, pp. 1551–1563, 1992. https://doi.org/10.1121/1.403896

[6] I. Setiawan, P. Murti, W. N. Achmadin, A. B. S. Utomo, and M. Nohtomi, “Design, construction and evaluation of a standing wave thermoacoustic prime mover,” in AIP Conference Proceedings, 2016, vol. 1717. https://doi.org/10.1063/1.4943482

[7] T. Jin, B. Zhang, K. Tang, R. Bao, and G. Chen, “Experimental observation on a small-scale thermoacoustic prime mover,” J. Zhejiang Univ. A, vol. 8, no. 2, pp. 205–209, 2007. https://doi.org/10.1631/jzus.2007.a0205

[8] K. Tang, G. B. Chen, T. Jin, R. Bao, B. Kong, and L. M. Qiu, “Influence of resonance tube length on performance of thermoacoustically driven pulse tube refrigerator,” Cryogenics (Guildf)., vol. 45, no. 3, pp. 185–191, 2005. https://doi.org/10.1016/j.cryogenics.2004.10.002

[9] N. M. Hariharan, P. Sivashanmugam, and S. Kasthurirengan, “Influence of stack geometry and resonator length on the performance of thermoacoustic engine,” Appl. Acoust., vol. 73, no. 10, pp. 1052–1058, 2012. https://doi.org/10.1016/j.apacoust.2012.05.003

[10] A. C. Alcock, L. K. Tartibu, and T. C. Jen, “Experimental investigation of an adjustable standing wave thermoacoustic engine,” Heat Mass Transf. und Stoffuebertragung, vol. 55, no. 3, pp. 877–890, 2018. https://doi.org/10.1007/s00231-018-2469-1

[11] G. Zhou, Q. Li, Z. Li, and Q. Li, “Influence of resonator diameter on a miniature thermoacoustic Stirling heat engine,” Chinese Sci. Bull., vol. 53, no. 1, pp. 145–154, 2008. https://doi.org/10.1007/s11434-008-0019-9

[12] G. . Swift, “Thermoacoustics : A Unifying Perspective for Some Engines and Refrigerators,” Leb. und-Technologie, vol. 35, no. 7, p. 636, 2002. https://doi.org/10.1016/s0023-6438(02)90911-4

[13] Sugiyanto, S. Kamal, J. Waluyo, and A. Widyaparaga, “Simulation of Close-Open Standing Wave Thermoacoustic Engine Toward Variation of Resonator Diameter,” in 4th International Conference on Science and Technology, ICST 2018, 2018, vol. 1, pp. 1–5. https://doi.org/10.1109/ICSTC.2018.8528657

[14] B. Ward, J. Clark, and G. Swift, Design environment for low-amplitude thermoacoustic energy conversion (DeltaEC), vol. 122, no. 5. 2007. https://doi.org/10.1121/1.2942768

[15] T. D. Rossing, Springer Handbook of Acoustic. LLC New York: Springer Science+Business Media, 2007.

[16] T. Biwa, Y. Tashiro, H. Nomura, Y. Ueda, and T. Yazaki, “Experimental verification of a two-sensor acoustic intensity measurement in lossy ducts,” J. Acoust. Soc. Am., vol. 124, no. 3, pp. 1584–1590, 2008. https://doi.org/10.1121/1.2953311

[17] Y. Takeyama, S. ichi Sakamoto, and Y. Watanabe, “Study on the setting position of a prime mover in the coaxial-type thermoacoustic cooling system: Comparison with the straight-tube-type thermoacoustic system,” Jpn. J. Appl. Phys., vol. 57, no. 7, 2018. https://doi.org/10.7567/JJAP.57.07LE14

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 306-313 (2019)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Priority Incorporated Zone Based Distributed Clustering Algorithm for Heterogeneous Wireless Sensor Network

Hamid Ali Abed AL-Asadi [*,1,2], Abdulhadi Alhassani[1] , Nor Azura Ahmed Hambali[3], Mustafa Abdulazeez AlSibahee[1,4], Saif Ali Alwazzan[1], Ali Mohammed Jasim[1]

[1]*Communications Engineering Department, Iraq University College, Basra, 61004, Iraq.*

[2]*Computer Science Department, Basra University, Basra, 61004, Iraq*

[3]*Semiconductor Photonics & Integrated Lightwave Systems (SPIILS), School of Microelectronic Engineering, Universiti Malaysia Perlis, Pauh Putra Main Campus, 02600 Arau, Perlis, Malaysia*

[4]*Shenzhen Institute of Huazhong University of Science and Technology, Shenzhen, China*

ARTICLE INFO

Article history:
Received: 29 July, 2019
Accepted: 21 September, 2019
Online: 15 October, 2019

Keywords:
Heterogeneous wireless sensor network
Priority
Zone based clustering
Distributed clustering
Supporting cluster head
Clustering
Sensor nodes

ABSTRACT

Wireless sensor networks (WSNs) are considered to be the currently flourishing scientific domain, thereby found to be applicable in numerous industrial and domestic applications. As per the mathematical results in Pulse-coupled oscillator (PCO), it has been predicted that, numerous iterations are needed for convergence, leading to increased power consumption. Biologically inspired solutions are greatly applicable for recovering coverage issues and efficient routing processes. In Hybrid energy efficient distributed clustering (HEED), to find a node with lowest communication cost, large number of iterations is needed, thereby leading to larger time duration for finding such node, and this is considered as the prevalent drawback, resulting in significant power consumption. In Optical low energy adaptive clustering hierarchy (O-LEACH), the cluster head selection is based on randomness, resulting in easy cluster failure. Hence, for cases, where power minimization and higher network lifetime is to be achieved to a larger extent, the existing strategies shall not be applicable, due to few restrictions. In this paper, a priority incorporated zone based distributed clustering algorithm, the Better Integrated and Optimized Low Energy Adaptive Clustering Hierarchy (BIO-LEACH) has been proposed for heterogeneous WSN. The methodology of this distributed clustering algorithm possesses three distinct features. First, the given clustering area will be divided in to different clusters and each cluster will be assigned with priority. The cluster which is highly sensitive and which needs frequent data recording will be given highest priority. The clusters in which the priority is assigned, takes multiple sensing and communication even in one cycle. But, the clusters where priority is not assigned, only one sensing and data transmission will be allowed for one cycle. Second, the clusters possessing priorities will have one cluster head (CH) and two supporting cluster head (SCH), but the clusters that do not have priorities will have only one cluster head. Third, the clusters possessing priorities will be possessing cluster nodes more than that of the clusters without priorities, so as to avoid cluster failure. Simulation results have been done to evaluate the performance of the proposed algorithm in terms of number of cluster head selection, amount of energy consumed and number packets received.

[*]Corresponding Author Hamid A. AL-Asadi, Head of Communications Engineering Department, Iraq University College, Basra, 61004, Iraq. Computer Science Department, Basra University, Basra, 61004, Iraq. Email: hamid.alasadi@iuc.edu.iq, hamid.abed@uobasrah.edu.iq

1. Introduction

In a typical wireless sensor networks (WSN), there will be hundreds to thousand number of sensor nodes that will be capable of functioning in unattended environments [1]. The sensor nodes in such networks collect information from the predestined environment and communicate with each other in a cooperative manner. These wireless sensor networks form the basic functioning module of Internet of Things (IoT) [2], thereby sensed data can be transmitted over the IoT, for numerous real-world applications like healthcare and military. These sensor networks therefore form the basic research focus in the present era, and also when these sensor nodes function as actuators, they could be intelligent to control a particular situation based on the sensed values or reading. These actuator applications are quite commonly applicable in military applications, earthquake detection and alerting, environmental monitoring applications, etc. Considering a single sensor node, it possesses a sensing unit, microcontroller unit, storage unit, radio communication unit and an antenna [3-5]. But, these sensor nodes are tiny in size, thereby power becomes the main area to be concentrated. Unlike other networks, sensor networks are mostly preferred to function in unattended environments where even human live is merely possible, thereby the power consumption of the batteries have to be clearly controlled. Therefore, bringing out an energy efficient strategy to minimize the power consumption of the sensor nodes has become the main view of recent researchers. One such method to attain these functionalities is by effectively forming clusters.

These sensor networks can be employed to measure single/multiple parameters using the same network, and hence heterogeneous WSN comes in to picture. The main difference between homogenous and heterogeneous wireless sensor network, in one point of view is that, the former uses identical sensor throughout the sensor field, but the latter uses different types of sensors within the sensor field.

Therefore, in a heterogeneous WSN, the main target will be towards reducing the power consumption of the sensor nodes and prolonging the overall lifetime of the network. Heterogeneous WSN possesses a group of sensor nodes that are different in functionalities, thereby deploying them, and topology management is also a serious concern [6].

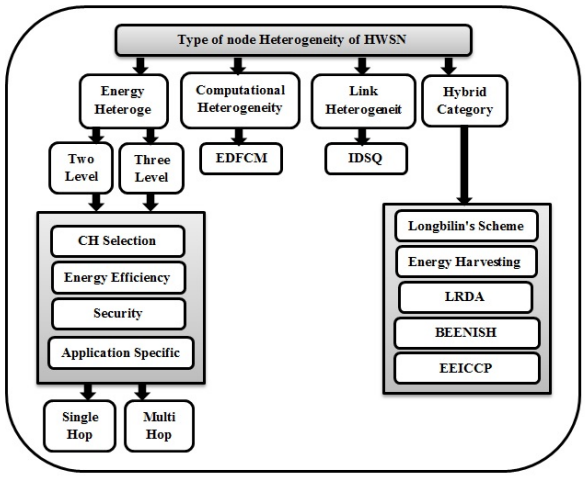


Figure 1: Taxonomy of heterogeneous schemes.

Fig.1 defines the taxonomy of heterogeneous protocols for WSNs. Heterogeneity of a sensor node may be based on energy, computational and link heterogeneities. Energy heterogeneity specifies the replacement capabilities of batteries in a sensor node, as the energy is the main constraint when compared to computational complexities and link failures in a heterogeneous WSN. Computational heterogeneity specifies that the particular sensor node possesses higher features like speed, storage, battery, etc., when compared to ordinary sensor nodes that too directly depends on energy heterogeneity. Processing highly complex data along with increased storage capabilities, forms the specific feature of a heterogeneous wireless sensor network. The communication distance between two sensor nodes is mainly considered as the parameter for link heterogeneity. When we go for long distance communication, communication bandwidth also forms a major concern in a heterogeneous WSN.

This paper has been structured as follows. An introduction to various concepts in a wireless sensor network, mainly focussing on heterogeneity, has been discussed in Section 1. The literature survey of several algorithms towards the closer vicinity of this research work is discussed in Section 2. The features of the proposed BIO-LEACH algorithm, and how it differs from the traditional algorithms has been clearly discussed in Section 3. Simulation results along with necessary discussions have been elaborated in Section 4. Finally, Section 5 concludes the paper.

2. Literature Review

Few related researches work carried out by researchers in the closer vicinity of our research work is discussed here, that includes, Pulse-coupled oscillator (PCO), Biologically inspired solutions, Hybrid energy efficient distributed clustering (HEED) and Optical low energy adaptive clustering hierarchy (O-LEACH).

Pulse-Coupled Oscillator (PCO) Method: Based on PCO, the first proposal started for attaining synchronization in a heterogeneous WSN. But, prior to that numerous research works have been carried out with theoretical and mathematical models, but they did not concentrate on the hardware part. This hardware oriented PCO strategy mainly concentrated on the physical layer parameters and employed point to point communication topology. But, subsequently many approaches have been proposed based on PCO for ultra wideband (UWB) communication applications [7]. This PCO falls under the category of energy efficient synchronization protocols for heterogeneous WSN. As per the mathematical results, it has been predicted that numerous iterations are needed for convergence. The mapping is done in continuous manner so as to attain convergence in these sensor networks. Even though, higher efforts were taken towards convergence and synchronization, these strategies will not guarantee effective convergence.

Biologically Inspired Solutions: These biologically inspired solutions are typically designed so as to imitate the basic functionalities of naturally inspired metaheuristic. Apart from attaining energy efficiency and prolonging network lifetime, these biologically inspired solutions offer numerous interesting benefits in WSN [8-10]. These strategies are normally considered as an effective tool that are generally simpler, highly scalable and offers distributed approaches for heterogeneous WSNs. These strategies are normally considered as the replica of optimizing real time

problems. Considering few examples, the ant colony optimization strategy serves as an effective tool for multipath clustering and routing, particle swarm optimization is considered as an effective approach for generating optimal number of clusters, the genetic algorithm is mainly employed to recover the coverage issues, and artificial bee colony algorithm mainly results in formation of efficient clustering and routing processes. These approaches are found to be applicable for controlling automated smart lighting systems and also in photovoltaics [11-14].

Hybrid Energy Efficient Distributed Clustering (HEED): This particular approach is mainly formulated so as to attain effective clusters in a wireless sensor network [15]. HEED algorithm mainly depends upon residual energy and communication cost for forming effective clusters. In case of Optical Low Energy Adaptive Clustering Hierarchy (O-LEACH), the cluster head is selected in random manner, but HEED was proposed for eradication of random cluster head (CH) selection. Initialization, repetition and finalization phases are the three different phases carried out while forming clusters in WSN based on HEED. The initial probability of a node to become CH for the current round is decided in the initialization phase of HEED. The node with lowest communication cost will be elected as CH in the repetition phase of HEED. And, in finalization phase, the election of cluster head will however be properly settled. In order to find a node with lowest communication cost, numerous iterations are needed, thereby leads to larger time duration for finding such node and this is considered as the prevalent drawback in HEED.

Optical Low Energy Adaptive Clustering Hierarchy (O-LEACH): This strategy was proposed to connect two separate and longer distanced sensor networks using an optical and distributed fiber link, and every other clustering process is followed as per the traditional LEACH algorithm. These two wireless sensor nodes are considered to possess huge number of wireless sensor nodes, and based on the nature of required applications, it could be assigned that these sensors shall communicate or shall not communicate with other sensor nodes. This optical fiber serves as the connectivity between two separate sensor fields, therefore at the edge of the network, base station (BS) will be positioned. The CH performs data aggregation process so as to further reduce the quantity of data that needs to be communicated to the base station [16-17]. As per the traditional LEACH concept, the cluster head selection is carried out in a random manner, and this is mainly considered as the common drawback during the cluster formation in O-LEACH algorithm also. The aggregated data is forwarded from the cluster head to the base station either by direct forward or hopping through other cluster heads.

2.1. Problem Statement

Considering heterogeneous network, as per the mathematical results in PCO, it has been predicted that, numerous iterations are needed for convergence, leading to increased power consumption. Hence, for cases, where power minimization and higher network lifetime is to be achieved, this strategy shall not be applicable. Biologically inspired solutions are greatly applicable for recovering coverage issues and efficient routing processes. In HEED, to find a node with lowest communication cost, large number of iterations is needed, thereby leading to larger time duration for finding such node, and this is considered as the prevalent drawback, resulting in significant power consumption. In O-LEACH, the cluster head selection is based on randomness, resulting in easy cluster failure. Hence, to overcome these issues, formulation of a novel clustering strategy, which increases the network lifetime and decreases the power consumption, is essential.

3. Proposed BIO-LEACH Algorithm

3.1. Methodology

The proposed distributed clustering algorithm possesses three distinct features. First, the given clustering area will be divided in to different clusters and each cluster will be assigned with priority, which is not observed in O-LEACH. The cluster which is highly sensitive and which needs frequent data recording will be given highest priority. The clusters in which the priority is assigned, takes multiple sensing and communication even in one cycle. But, the clusters where priority is not assigned, only one sensing and data transmission will be allowed for one cycle. Second, the clusters possessing priorities will have one cluster head (CH) and two supporting cluster head (SCH), but the clusters that do not have priorities will have only one cluster head. Third, the clusters possessing priorities will be possessing cluster nodes more than that of the clusters without priorities, so as to avoid cluster failure. In case of O-LEACH, it was proposed to connect two separate and longer distanced sensor networks using an optical and distributed fiber link, and every other clustering process is followed as per the traditional LEACH algorithm. The main difference between the 0-LEACH and BIO-LEACH is that, clustering process in O-LEACH is as per the traditional LEACH algorithm, which results in reduced network lifetime, but, the proposed algorithm employs aforementioned novelties in clustering process, thereby results in improved network lifetime. Fig.2 shows the basic articulation of the BIO-LEACH algorithm.

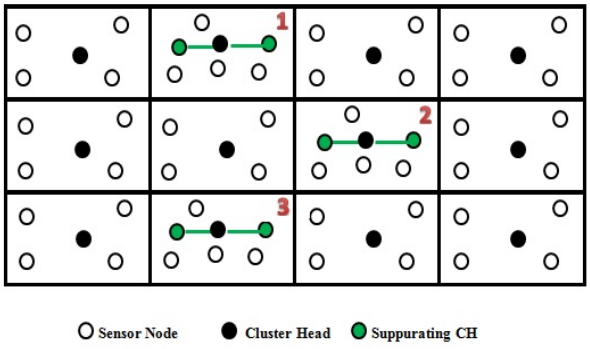


Figure 2: Basic articulation of BIO-LEACH algorithm.

The clustering field scenario considered for the proposed scheme is divided into 12 clusters of equal dimensions. Clusters 2, 7 and 10 are only assigned with priorities, but the other clusters are considered as normal clusters. The clusters in which the priorities are assigned has one cluster head and two supporting cluster head (SCH). Also, the sensor nodes are higher in number when compared with the ordinary clusters. The main reason we are assigning a greater number of sensor nodes is that, these priority clusters will do more sensing and data communication when compared to the ordinary clusters. The operational flowchart of the proposed algorithm is depicted in Fig.3. The three main phases in BIO-LEACH are cluster formation, Data transmission and CH-CH communication.

3.2. Assignment of Priority and Synchronization

In the proposed strategy, it is supposed for every node to possess a clock which varies from 0 to T, and this could be expressed mathematically as,

$$\phi_i(t) = \frac{t}{T}\phi_i(0) \mod 1 \quad (1)$$

Where, I represent a particular node and $\phi_i(0)$ signify the offset period. Equation 1 clearly specifies that the state of a particular node at $t > 0$, therefore the sensed information correlates with the firing time. The sensor node i that hears the advertisement of its neighbouring nodes will update its clock based on the function given below,

$$\phi_j(t^+) = \min\{f^{-1}(f(\phi_j(t)) + \varepsilon), 1\} \quad (2)$$

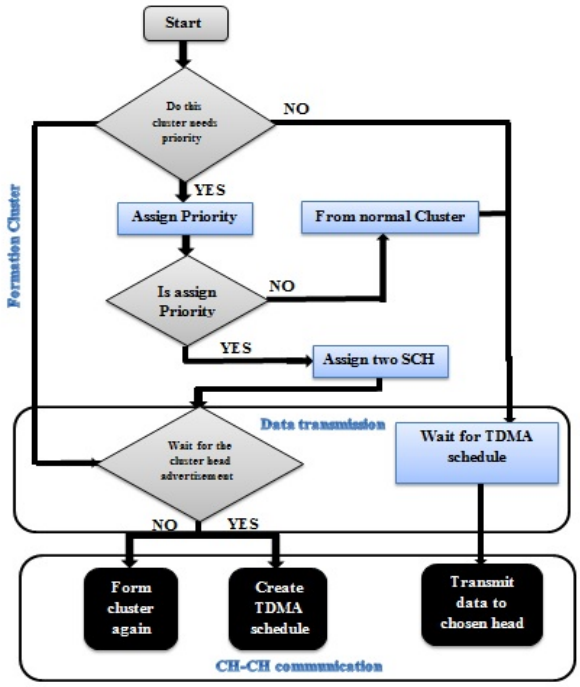


Figure 3: Operational flowchart of the Proposed Algorithm.

Where $\varepsilon > 0$ represents the coupling strength and $f(.)$ represents the predestined sensor function such that $f(0) = 0, f(1) = 1 \; and \; f(x) > 0; \; \forall x \in (0,1)$ A typical example correlating the state x_j and phase ϕ_j can be integrated as a function given below,

$$x_j = f(\phi_j) = \frac{1}{b}\log\left(1 + \left[e^b - 1\right]\phi_j\right) \quad (3)$$

A particular node j that hears the advertisement message shall be expressed and correlated mathematically as,

$$\phi_j(t^+) = \min\{a_1\phi_j(t) + a_0, 1\} \quad (4)$$

$$a_1 = e^{b\varepsilon} \quad (5)$$

$$a_0 = \frac{e^{b\varepsilon} - 1}{e^b - 1} \quad (6)$$

Thus, the condition for an updated node j to get synchronized with the original node i shall be integrated and expressed in mathematical form as below,

$$\phi_j(t^+) = \min\{f^{-1}(f(\phi_j(t)) + \varepsilon), 1\} \quad (7)$$

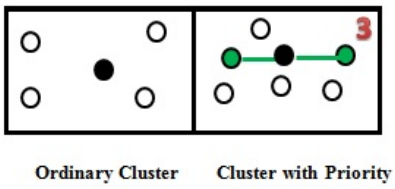


Figure 4: Cluster Formation Scenario.

Two different clusters are formed: one with priority and other normal clusters. Fig.4 demonstrates how different types of clusters are formed in the proposed strategy. User can assign manually as per Equations (4)-(6), how many clusters needs priority and how many which do not need it, based on the requirements. Equations (1)-(7) describes the assignment of priority in the proposed algorithm, thereby forms an essential role when compared with O-LEACH, in overcoming the drawbacks of O-LEACH.

3.3. Data Transmission from CH to the BS

Fig.5 shows the data transmission path in the proposed approach. The cluster nodes will sense the data and keep it ready, once the TDMA time slot is assigned by the cluster head (CH), these nodes forward them to the cluster head. The main difference between the existing algorithms and the proposed approach is that, the existing algorithms do not use relay nodes.

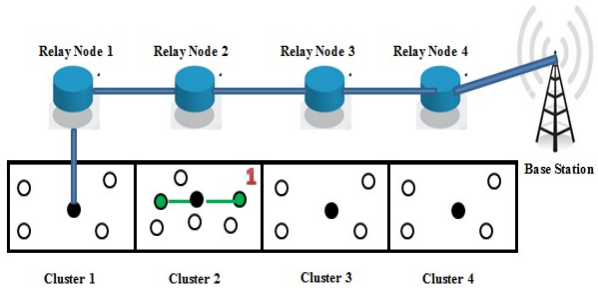


Figure 5:Data transmission path in the proposed approach.

Relay nodes are the nodes that are rich in resources, and in the proposed approach, these relay nodes are used only for communication between the CH and base station (BS). Relay nodes are employed between cluster head and the base station. These nodes carry the aggregated data from the cluster head to the base station. In most of the existing approaches, CHs sends the aggregated data to the base station, either directly or forwarding them through other cluster heads. This leads to sudden failure of cluster head. This shortcoming is avoided in the proposed approach, by the employment of relay nodes. Data transmission path is a multihop communication from CH to BS.

4. Results and Discussions

For simulation purpose, a square sensor field has been considered with N distinct sensor nodes with MxM dimension. The quantity of sensor nodes in the whole sensor field shall be expressed mathematically as,

$$S = \{s_1, s_2, ..., s_N\} \quad (8)$$

The sensor nodes are assumed to be in immobile fashion, thereby each and every sensor nodes are pre-assigned with one unique identifier for distinguishing the data collected by the nodes in the sensor field, and can be expressed mathematically as,

$$i \in [1, N] \quad or \quad 1 \le i \le N \quad (9)$$

We also have assumed that every sensor nodes are heterogeneous and possess variable functionalities, and also the batteries of the sensor nodes are capable to recharge based on some mechanisms.

Table 1: The used Notations in BIO-LEACH Algorithm

S.No	Notation	Description
1	Cl_m	Cluster Members
2	$CH_f(j)$	Followers of CH i
3	CH_c	CH Candidate
4	$CH_c(i)$	Set of CH_c received by sensor i
5	ε_l	Minimum Energy Level
6	ε_t	Quality Assessment Threshold
7	ε_d	Decay time
8	R_c	Communication Range
9	T_{-1}	Time of initialization step
10	T_{-2}	Time of CH Election step

The position $(x_i, y_i) \in R$ of a sensor node i is signified and considered as θ^i. Also, it is considered that, the sensor nodes do not know the exact position of other sensor nodes, but the cluster head and the sensor nodes are aware of the position of the base station. The BS is outside the sensors square, near the middle edge of the square, and has sufficient resources. Each sensor can receive a message from another, if it is in the communication range of the sender sensor. The communication distance corresponding to the sensing distance ds is expressed mathematically as,

$$R_c = 2.\pi \times d_s \quad (10)$$

For effective cluster formation in the proposed BIO-LEACH algorithm, the following parameters and notations are considered as represented in Table 1. Every simulation work has been done in NS2 for maximum of 100 rounds. Unlike O-LEACH and HEED approaches, the subsequent cluster heads are elected only when the threshold values of the prevailing cluster head become less than 1. This concept is not available in most of the existing distributed clustering algorithms. Subsequent to initialization step, the election of cluster heads wills commences at particular time T1. But, in case of next time duration T2, the sensor node that has the possibility to become a cluster head will send CHFOLLOW message. In circumstance when the sensor node acquires this candidature, it will estimate the hop length and hop count of each and every CH. In case, when the cluster head do not receive the CHFOLLOW message, it will change its status from node j to normal node i. In case of steady state phase, the data gets aggregated by following TDMA slots. The communication between CH and BS is accompanied with the help of CDMA, but communication in between local clusters will be done by TDMA.

Fig.6 shows the number of CHs and number of CH elections in BIO-LEACH during 100 runs. In the first 10 runs out of the 10 elected cluster head only 5 is finalized. In 20 runs, 6 get finalized out of 10 elections. In next stage, 5 get finalized out of 14 elections. Finally, in 100 runs, 4 get finalized out of 10. At an average, 45.63% represents the average percentage of cluster head finalized out of those selected initially. This figure is a decent value, which specifies the novel strategies followed for cluster head selection in the proposed approach. The clusters where priority is not assigned, only one sensing and data transmission will be allowed for one cycle. The clusters possessing priorities will have one cluster head and two supporting cluster head, but the clusters that do not have priorities will have only one cluster head. This concept helps in proper rotation of CHs when compared with the existing approaches, thereby aids in better CH selection.

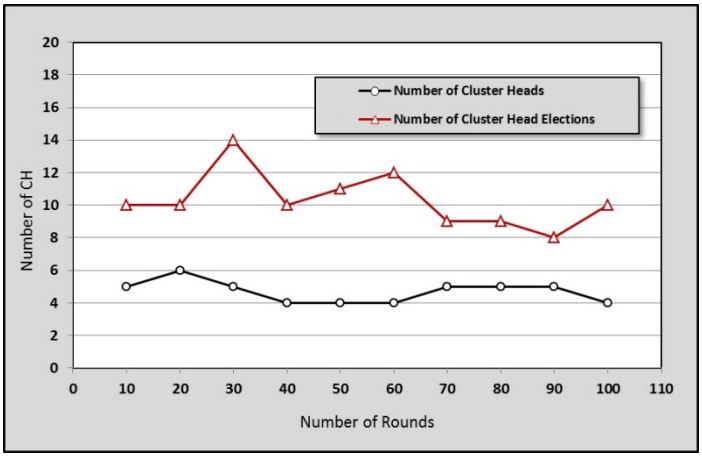


Figure 6: CH Election strategy outcomes.

Fig.7 specifies the network lifetime evaluation in BIO-LEACH. Network lifetime is nothing but the total span of period over which the sensor network will be alive. Also, in one angle, it could be specified as the time span for the first node to die, because the death of the first node implicitly indicates that there will be some loss in overall function of the sensor network. It also represents the case when, some network nodes die out of energy, and in this case the other network nodes being used to acquire the information from the dead node and transfer it to the base station. Here, initially for 1000 received packets, the number of nodes alive in O-LEACH, PSO and HEED are same as that of the proposed approach.

But, for 6000 received packets, the number of nodes alive in O-LEACH, PSO and HEED are 100, 400 and 450, but the proposed algorithm has a good number of alive nodes as 500.

Similarly, for 8000 received packets, the number of nodes alive in O-LEACH, PSO and HEED are 100, 200 and 350, but the proposed algorithm has a good number of alive nodes as 400. Finally, for 10000 received packets, the number of nodes alive in O-LEACH, PSO and HEED are 100, 100 and 100, but the proposed algorithm has a good number of alive nodes as 200. Thus, the loss of nodes is greatly prevented in the proposed approach. The average number of nodes alive in O-LEACH, PSO, HEED and BIO-LEACH are 2275, 3550, 4075 and 4325 . BIO-LEACH algorithm shows 47.39% improvement in number of nodes alive when compared to O-LEACH, 17.91% improvement in number of nodes alive when compared to PSO and 5.78% improvement in number of nodes alive when compared to HEED, respectively. This is mainly achieved by the incorporation of relay nodes, priority and clustering concepts.

Here, the clusters possessing priorities will be possessing cluster nodes more than that of the clusters without priorities, so as to avoid cluster failure. In case of O-LEACH, it was proposed to connect two separate and longer distanced sensor networks using an optical and distributed fiber link, and every other clustering process is followed as per the traditional LEACH algorithm. Thus, by the concept of using priority, the proposed strategy attains betterment in network lifetime.

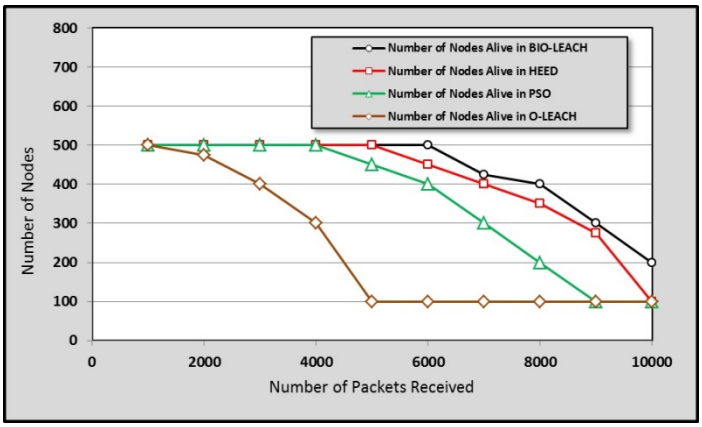


Figure 7: Network lifetime evaluation in BIO-LEACH.

Fig.8 describes the evaluation of amount of energy consumed in all the four approaches in terms or residual energies. Models proposed and implemented by Heinzelman normally attain better energy efficiency that implicitly signifies excellent power controlling and there is no any channel fading to happen. Initially for 1000 received packets, the residual energy in O-LEACH, PSO and HEED is 10 Joules, same as that of the proposed approach. But, for 6000 received packets, the residual energy in O-LEACH, PSO and HEED are 2 Joules, 4 Joules and 6 Joules, but the proposed algorithm has better residual energy as 5.5 Joules. Similarly, for 8000 received packets, the residual energy in O-LEACH, PSO and HEED are 0.5 Joules, 2 Joules and 3 Joules, but the proposed algorithm has a better residual energy as 4 Joules. Finally, for 10000 received packets, the residual energy in O-LEACH, PSO and HEED are 0.5 Joules, 1 Joule and 3 Joules, but the proposed algorithm has a good residual energy as 4 Joules. The average residual energy in O-LEACH, PSO, HEED and BIO-LEACH are 38.50 Joules, 49 Joules, 60.50 Joules and 64.50 Joules. BIO-LEACH algorithm shows 40.31% betterment in residual energy when compared to O-LEACH, 23.87% improvement in residual energy when compared to PSO and 6.20% improvement in residual energy when compared to HEED, respectively. For highly stable clusters, the residual graph should be linear with lesser irregularities, and from the graph it is evident that the proposed approach is highly stable.

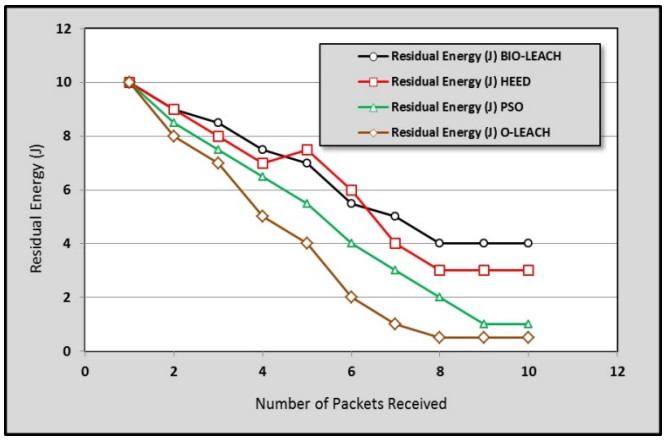


Figure 8: Evaluation of Amount of energy consumed.

Fig.9 describes the amount of packets received by the BS in all the four approaches. Congestion in wireless sensor network occurs because of loss or drop of packets, that mainly affects the network lifetime and throughput in these wireless sensor networks. Also, losses shall happen when the implemented network possess very poor security mechanisms. Initially for 10 rounds, the amount of packets received by the BS in O-LEACH, PSO and HEED is 10000, same as that of the proposed approach. But, for 40 rounds, the amount of packets received by the BS in O-LEACH, PSO and HEED are 6600, 6900 and 6400, but the proposed algorithm has better amount of packets received by the BS as 8800. Similarly, for 80 rounds, the amount of packets received by the BS in O-LEACH, PSO and HEED are 3900, 4200 and 3700, but the proposed algorithm has a better amount of packets received by the BS as 4 Joules. Relay nodes are employed between cluster head and the base station. These nodes carry the aggregated data from the cluster head to the base station. In most of the existing approaches, CHs sends the aggregated data to the base station, either directly or forwarding them through other cluster heads. This leads to sudden failure of cluster head. This shortcoming is avoided in the proposed approach, by the employment of relay nodes, thereby leading to number of successfully delivered packets. Also, this results in the decrease in communication delay, as seen in Fig.10.

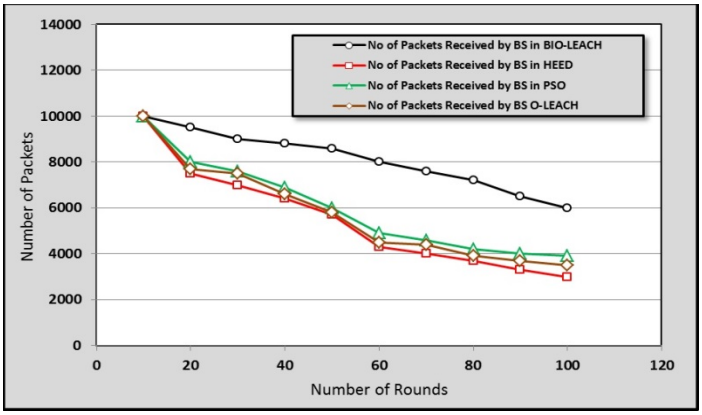


Figure 9: Amount of packets received by the BS in the Four Approaches.

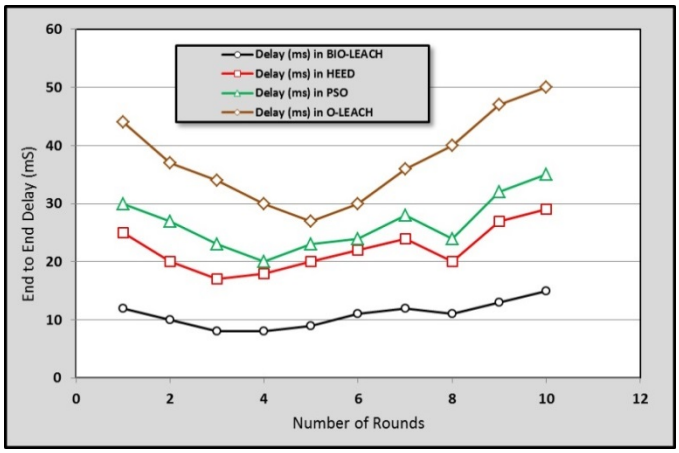


Figure 10: Average end-to-end delay during 100 runs for all the four Strategies.

Finally, for 100 rounds, the residual energy in O-LEACH, PSO and HEED are 3500, 3900 and 3000, but the proposed algorithm has a good amount of packets received by the BS as 6000. The average amount of packets received by the BS in O-LEACH, PSO, HEED and BIO-LEACH are 57600, 60100, 54900 and 81200. BIO-LEACH algorithm shows 29.06% betterment in amount of packets received by the BS when compared to O-LEACH, 25.98% improvement in amount of packets received by the BS when compared to PSO and 26.84% improvement in amount of packets received by the BS when compared to HEED, respectively. The slope of the curve is also linear with less irregularities, making it more stable.

Fig.10 describes the average end-to-end delay during 100 runs for all the four Strategies. End-to-end delay or one-way delay signifies the overall time duration expended by the packets for reaching the base station from the cluster node or cluster head. It is a communal word in IP network monitoring, and varies from round trip time, where the path length in one way direction is only considered. Here, initially for 10 runs, the average end-to-end delay in O-LEACH, PSO and HEED is 44ms, 30ms and 25ms, but for the proposed approach the average end-to-end delay is 12ms. But, for 40 runs, the average end-to-end delay in O-LEACH, PSO and HEED are 30ms, 20ms and 18ms, but the proposed algorithm has reduced average end-to-end delay as 8.1ms. Similarly, for 80 runs, the average end-to-end delay in O-LEACH, PSO and HEED are 40ms, 24ms and 20ms, but the proposed algorithm has a low average end-to-end delay as 11ms. Finally, for 100 runs, the average end-to-end delay in O-LEACH, PSO and HEED are 50ms, 35ms and 29ms, but the proposed algorithm has a reduced average end-to-end delay as 15ms. The average end-to-end delay in O-LEACH, PSO, HEED and BIO-LEACH are 375ms, 266ms, 222ms and 109.1ms.

5. Conclusion

In this paper, a priority incorporated zone based distributed clustering algorithm, the Better Integrated and Optimized Low Energy Adaptive Clustering Hierarchy (BIO-LEACH) has been proposed for heterogeneous wireless sensor networks. This distributed clustering algorithm possesses three distinct features. First, the given clustering area will be divided in to different clusters and each cluster will be assigned with priority. The cluster which is highly sensitive and which needs frequent data recording will be given highest priority. The clusters in which the priority is assigned, takes multiple sensing and communication even in one cycle. But, the clusters where priority is not assigned, only one sensing and data transmission will be allowed for one cycle. Second, the clusters possessing priorities will have one cluster head (CH) and two supporting cluster head (SCH), but the clusters that do not have priorities will have only one cluster head. Third, the clusters possessing priorities will be possessing cluster nodes more than that of the clusters without priorities, so as to avoid cluster failure. From the simulation results, BIO-LEACH algorithm shows 47.39% improvement in number of nodes alive when compared to O-LEACH, 17.91% improvement in number of nodes alive when compared to PSO and 5.78% improvement in number of nodes alive when compared to HEED, respectively. Also, the proposed strategy shows 40.31% betterment in residual energy when compared to O-LEACH, 23.87% improvement in residual energy when compared to PSO and 6.20% improvement in residual energy when compared to HEED, respectively. Moreover, the novel algorithm shows 29.06% betterment in amount of packets received by the BS when compared to O-LEACH, 25.98% improvement in amount of packets received by the BS when compared to PSO and 26.84% improvement in amount of packets received by the BS when compared to HEED, respectively. Hence this proposed algorithm exhibits highly stable clusters.

References

[1] C. Han, Q. Lin, J. Guo, L. Sun and Z. Tao. "A Clustering Algorithm for Heterogeneous Wireless Sensor Networks Based on Solar Energy Supply", MDPI Electronics. 7(103), 1-20, 2018. https://doi.org/10.3390/electronics7070103

[2] R. Hou, L. He ; S. Hu ; J. LuoL., "Energy-Balanced Unequal Layering Clustering in Underwater Acoustic Sensor Networks", IEEE Access. (6). 39685 – 39691, 2018. DOI: 10.1109/ACCESS.2018.2854276

[3] J. Corn, J. W. Bruce, "Clustering algorithm for improved network lifetime of mobile wireless sensor networks", IEEE International Conference on Computing, Networking and Communications (ICNC), 26-29, 2017. DOI: 10.1109/ICCNC.2017.7876282

[4] J. Yu, Y. Chen, L. Ma, B. Huang, and X. Cheng. "On connected target k-coverage in heterogeneous wireless sensor networks. Sensors", 16(1), 104, 2016. DOI: 10.3390/s16010104

[5] A. Sharma and J. Thakur. "An energy efficient network life time enhancement proposed clustering algorithm for wireless sensor networks", In International Conference of Parallel and Distributed Computing Systems (PDCS 2014), 111- 115. 2014. Available: http://hpc-ua.org/pdcs-14/files/proceedings/17.pdf

[6] A.Ari, B. Yenke, N. Labraoui, I. Damakoa, and A.Gueroui. "A power efficient cluster-based routing algorithm for wireless sensor networks: Honeybees Swarm Intelligence based approach", Journal of Network and Computer Applications, 69:77-97, 2016. DOI: 10.1016/j.jnca.2016.04.020

[7] X. Cai, Y. Duan, Y. He, J. Yang, and C. Li. "Bee-Sensor-C: An energy-efficient and scalable multipath routing protocol for wireless sensor networks", International Journal of Distributed Sensor Networks, 0-14, 2015. https://doi.org/10.1155/2015/976127

[8] R. Elhabyan and M. Yagoub. "PSO-HC: Particle swarm optimization protocol for hierarchical clustering in wireless sensor networks", In Collaborative Computing: Networking, Applications and Work-sharing (CollaborateCom), IEEE International Conference on, 417-424, 2014. DOI:10.4108/icst.collaboratecom.2014.257336

[9] R. de Friffein, R. Loomba, and B. Jennings. "Selecting energy efficient cluster-head trajectories for collaborative mobile sensing", In IEEE Global Communications Conference (GLOBECOM), 1-7. 2015. DOI:10.1109/glocom.2014.7417727

[10] Z. Al-Mekhlafi, Z. Hanapi, M. Othman, and Z. Zukarnain. "Travelling wave pulse coupled oscillator (TWPCO) using a self-organizing scheme for energy-efficient wireless sensor networks: PLoS One", 12(1), 01-27, 2017. https://doi.org/10.1371/journal.pone.0167423

[11] S. Shamshirband, N. Anuar, M. Kiah, V. Rohani, D. Petkovic, S. Misra, and A. Khan. "Co-FAIS: Cooperative fuzzy artificial immune system for detecting intrusion in wireless sensor networks", Journal of Network and Computer Applications, 42:102(117), 2014. DOI:10.1016/j.jnca.2014.03.012

[12] L. Yan., W. Pan., B. Luo., X. Li. and J. Liu., "Modified Energy-Efficient Protocol for Wireless Sensor Networks in the Presence of Distributed Optical Fiber Sensor Link," IEEE Sensors Journal, Vol. 11, 1815-1819, 2011. DOI: 10.1109/JSEN.2010.2104317

[13] O. Younis, and S. Fahmy, , "HEED: A Hybrid Energy-Efficient Distributed Clustering Approach for Ad Hoc Sensor Networks," IEEE Transactions on Mobile Computing, 2004. DOI: 10.1109/TMC.2004.41

[14] Y. Guo, C. Xie, and Z. Jing, 'Improvement on LEACH by combining adaptive cluster head election and two-hop transmission', Proceedings of the 9th International Conference on Machine Learning and Cybernetics, 11-14, 2010. DOI: 10.1109/ICMLC.2010.5580988

[15] Z. Haitao, and B. Wenshao, 'A clustering routing protocol for energy balance of wireless sensor network based on simulated annealing and genetic algorithm', International Journal of Hybrid Information Technology, 7(2), 71-82, 2014. http://dx.doi.org/10.14257/ijhit.2014.7.2.08

[16] L. Han, "LEACH-HPR: An energy efficient routing algorithm for heterogeneous WSN", Proceedings of the IEEE International Conference on Intelligent Computing and Intelligent Systems, 507-511, 2010. DOI:10.1109/ICICISYS.2010.5658274

[17] J. Hill, R. Szewczyk, A. , S. Hollar, D. Culler, K. Pister, 'System architecture directions for networked sensors', Proceedings of the 9th International Conference on Architectural Support for Programming Languages and Operating Systems, 93-104, 2000. DOI:10.1145/356989.356998

ASTES

Advances in Science, Technology and Engineering Systems Journal
Vol. 4, No. 5, 314-320 (2019)
www.astesj.com

ASTES Journal
ISSN: 2415-6698

Stability Analysis of a Linear Model Predictive Control and its Application in a Water Recovery Process

Paolo Mercorelli*

Institute of Product and Process Innovation, Leuphana University of Lueneburg, Universitaetsallee 1, D-21339 Lueneburg, Germany

ARTICLE INFO

Article history:
Received: 17 February, 2019
Accepted: 13 September, 2019
Online: 25 October, 2019

Keywords:
Model Predictive Control
Stability Analysis
Applications

ABSTRACT

This paper examines application of linear general model predictive control (LGMPC). The stability of the LGMPC is proven by means of a demonstration of a theorem stating a sufficient and constructive condition. Lower bounds conditions are found for one of these matrices and then a system with saturation is taken into consideration. The conditions can be interpreted through physical aspects. The results obtained were tested by means of computer simulations and an example with a water recovery process is considered.

1 Introduction

Model predictive control (MPC) is a technique that can be used to calculate feedback control for both linear and nonlinear systems using online optimisation techniques. Due to its great flexibility, MPC is likely to be the method with the most practical applications among modern control algorithms. MPC is suitable for limited, multivariable systems and for control problems where regulation of the control function is very difficult or even impossible. One of the great advantages of MPC is its ability to handle the limitations of the system, which makes these methods very interesting to industry.

Linear control theory generally deals with systems that are defined in continuous time and that can be described by ordinary differential equations in which the process model is the cornerstone of MPC. The model should fully capture the process dynamics and also be set up so that the predictions can be calculated.

The application of MPC to mechatronic systems for servo design attracts the attention of many scientists because of the continuous development of microprocessor technology. Mechatronic systems such as electrical motor control [1], two-stage actuation system control and machine tool chattering control [2] have shown promising results. Various advanced techniques are being rapidly developed that integrate MPC to improve performance [3]. However, the sampling frequencies previously used in the research literature may be too low for general mechatronic systems. Moreover, simulations show that the existence of modelling errors leads to obvious steady-state errors. Both of these points, especially during real-time implementation, may influence the performance of mechatronic systems. For example, the calculation of the solution of MPC techniques in an off-line explicit way has been demonstrated [4, 5], and MPC has been also applied to piezoelectric actuators [6]. Finding the conditions for stability is one of the interesting issues in optimisation.

The goal of this paper is to find the lower bounds of a matrix characterising the cost function that gives stability in an optimal solution. The results obtained by means of a proportional integral (PI) controller by Mercorelli et al. [7] and Mercorelli [8, 9] are extended in this contribution. This work considers a water distillation system in which water is separated from mud and impurities using a standard solution based on a heating system combined with a pressure control structure. The experimental setup represents a good example in terms of sustainability. Nevertheless, the algorithm proposed can be applied to any other linear system and represents an advancement in terms of the choice of weighting parameters for the cost function.

The paper is organized as follows. Section 2 is dedicated to model a water recovery system. Section 3 presents backgrounds of MPC. In Section 4 and 5, the structure of LGMPC is analysed without and with input saturation respectively. In Section 6, the proposed control technique is applied to a simulated example in the context of a Water Recovery Process. Conclusion closes the paper.

2 Mathematical model of the system

Figure 1 shows a schematic of the system being considered. The system comprises three elements: a boiler, a compressor and an evaporator. The waste water is located in the middle of the boiler.

*Paolo Mercorelli, Institute of Product and Process Innovation, Leuphana University of Lueneburg, Universitaetsallee 1, D-21339 Lueneburg, Germany & mercorelli@uni.leuphana.de

The main nomenclature
$m_{in}(t)$: input mass flow (kg/sec)
$m_o(t)$: output mass flow (kg/sec)
$m(t)$: mass (Kg)
$\frac{dm(t)}{dt}$: mass flow (kg/sec)
$p(t)$: pressure inside the evaporator (Pa)
T: temperature (K)
V: volume (l)
p_i: initial pressure (Pa)
p_d: desired pressure (Pa)
R_g: vapour constant
A_w: anti-windup signal
K_b: weight factor for the anti-windup action
$V_{\mathcal{L}}$: a Lyapunov function
$\mathbf{A}_k$: discrete system matrix
$\mathbf{B}_k$: discrete input matrix
$\mathbf{H}_k$: output matrix
$\mathbf{G}_p$: model predictive control state matrix
$\mathbf{F}_{1p}$: model predictive control "delta" input matrix
$\mathbf{F}_{2p}$: model predictive control input matrix

A vapour chamber hosts water vapour produced by heating the waste water in the upper part. In the first phase, the resistor system heats the waste water. After that, vapour appears in the chamber. The heating process is then turned off and the compressor is turned on. The pressure in the vapour chamber must be reduced using the compressor with the help of a mass flow. This mass flow ($m_0(t)$) represents cleaned and distilled water in the output of the evaporator. To be more precise, the vapour is condensed and transfers heat to the waste water in the evaporator. The low pressure in the vapour chamber allows new waste water to penetrate into the boiler. In this case, the compressor plays the role of a controller, and the mass flow $m_0(t)$, with the constraint that $m_0(t) > 0$, represents its output. The compressor includes an asynchronous motor being regulated by means of an inverter, which is driven by a pulse-width modulated (PWM) signal that converts the output of the LMPC controller into a frequency. The error signal occurs in the input to the LMPC. The section concerning the simulation demonstrates the details of the control scheme. The dynamics of the asynchronous motor with the inverter and the other converters are not considered in this analysis, but they are faster than the dynamics of the controlled process.

In the first phase, the pressure in the container is approximately 1.013 bar. The pressure increases during heating as the water starts to evaporate. At a lower pressure, the water boils and evaporates faster. New water is added under control. The "internal heat exchanger" helps to condense the vapour, which now contains no impurities. The controller design must take the dynamic model of the system into account. As explained above, the following equations can be taken into account, considering that the process of regulation begins when the steam is in the boiler. The idealised model can be presented as follows:

$$\begin{aligned} \frac{dm(t)}{dt} &= m_{in}(t) - m_o(t) \\ \frac{dp(t)}{dt} &= \frac{dm(t)}{dt}\frac{R_g T}{V}, \end{aligned} \tag{1}$$

with $m_{in}(t)$ being a stepwise positive constant function.

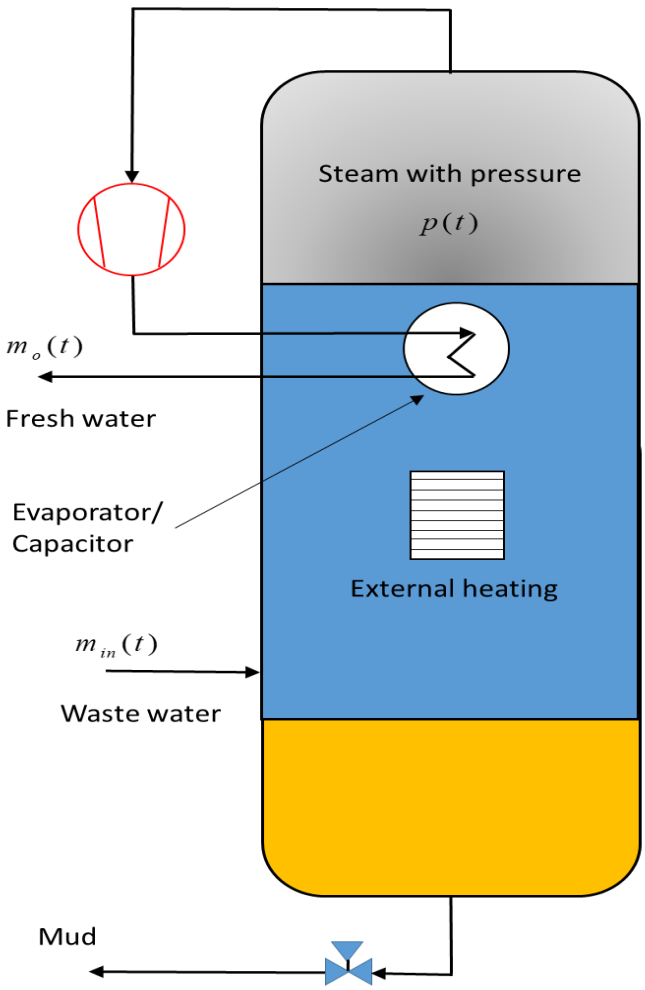


Figure 1: Boiler system

Considering the forward Euler discretisation with sampling time T_s, this expression is obtained:

$$\begin{aligned} m(k+1) &= T_s(m_{in}(k) - m_o(k)) + m(k) \\ p(k+1) &= (m_{in}(k) - m_o(k))\frac{R_g T}{V} + p(k), \end{aligned} \tag{2}$$

and thus

$$\underbrace{\begin{bmatrix} m(k+1) \\ p(k+1) \end{bmatrix}}_{\hat{\mathbf{z}}(k+1)} = \underbrace{\begin{bmatrix} T_s & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{A}_k} \underbrace{\begin{bmatrix} m(k) \\ p(k) \end{bmatrix}}_{\hat{\mathbf{z}}(k)} + \underbrace{\begin{bmatrix} T_s \\ \frac{T_s R_g T}{V} \end{bmatrix}}_{\mathbf{B}_k} \underbrace{(m_{in}(k) - m_0(k))}_{u_{mpc(t)}} \tag{3}$$

3 Model predictive control

The process model is the cornerstone of MPC. The model should fully capture the process dynamics and also be set up so that the predictions can be calculated. At the same time, it should be intuitive and allow a theoretical analysis. The process model is necessary

to calculate the predicted output quantities $y(t+k|t)$, also referred to as outputs, in a future instance, where $y(t+k|t)$ designates the output at time $t+k$ from time t. The various strategies of MPC can use numerous models to show the relationship between the output quantities and the measurable input quantities (inputs). A disturbance model can also be considered to describe behaviour not reflected by the process model, as well as non-measurable input magnitudes, measurement noise and model error. However, we will not consider these models at this point. Virtually any form of modelling in an MPC formulation can be used, but the following are the most common.

3.1 Problem definition and general idea of MPC

The problem of the model-predictive approach is a generalised form of stabilisation: the so-called tracking problem. The goal here is that the output quantities $y(t+k)$ follow a given reference trajectory $w(t+k)$, i.e. the deviation of the output quantities $y(t+k)$ from a known reference trajectory $w(t+k)$ should be minimised within a certain time window (horizon N) in order to keep the process as close as possible to this trajectory. This can be achieved by influencing the future output variables by a control sequence $u(t+k)$ to be calculated within a finite horizon N. For this purpose, a goal function J is set up, which is normally quadratic in form. The reference trajectory w is assumed to be known. The calculated output quantities $y(t+k)$, on the other hand, depend on the chosen model description and the future control signals $u(t+k)$ or the vector of the future control function changes $\Delta u(t) = u(t) - u(t-1)$. The basic idea of MPC can be explained in the following way. The future control quantities $y(t+k)$ at time t are to be predicted over a finite horizon N using the process model. These calculated output quantities depend on known values of the instance t and the future control signals $u(t+k)$, which are to be calculated and output to the system. The future control sequence is calculated by optimising a certain criterion, in most cases by minimising a target function J. Starting from the current time t, this objective function is now set up over the control horizon N and, with suitable optimisation methods, dependent on the zuk control values $u(t+k)$ and control value changes Δu are minimised.

3.2 The considered case

Just two samples of the model approach are considered:

$$\hat{\mathbf{z}}(k+1/k) = \mathbf{H}_k\mathbf{A}_k\mathbf{z}(k/k) + \mathbf{H}_k\mathbf{B}_k u_{mpc}(k). \tag{4}$$

If $\Delta u_{mpc}(k) = u_{mpc}(k) - u_{mpc}(k-1)$, then $u_{mpc}(k) = \Delta u_{mpc}(k) + u_{mpc}(k-1)$ and

$$\hat{y}_h(k+1) = \mathbf{H}_k\mathbf{A}_k\hat{\mathbf{z}}(k/k) + \mathbf{H}_k\mathbf{B}_k(\Delta u_{mpc}(k) + u_{mpc}(k-1)), \tag{5}$$

where matrix $\mathbf{H}_k$ selects the second state variable, and thus $y_h(t)$ represents the pressure.

$$\hat{y}(k+2) = \mathbf{H}_k\mathbf{A}_k^2\hat{\mathbf{z}}(k/k) + \mathbf{H}_k\mathbf{A}_k\mathbf{B}_k(\Delta u_{mpc}(k) + u_{mpc}(k-1)) + \mathbf{H}_k\mathbf{B}_k(\Delta u_{mpc}(k+1) + u_{mpc}(k)). \tag{6}$$

It must be shown that the following vectorial expression holds:

$$\hat{\mathbf{Y}}_h(k) = \mathbf{G}_p\mathbf{x}(k) + \mathbf{F}_{1p}\Delta\mathbf{U}_{mpc}(k) + \mathbf{F}_{2p}u_{mpc}(k-1), \tag{7}$$

where

$$\hat{\mathbf{Y}}_h(k) = \begin{bmatrix}\hat{y}_h(k+1)\\ \hat{y}_h(k+2)\end{bmatrix}, \qquad \Delta\mathbf{U}_{mpc}(k) = \begin{bmatrix}\Delta u_{mpc}(k)\\ \Delta u_{mpc}(k+1),\end{bmatrix} \tag{8}$$

and matrices $\mathbf{G}_p$, $\mathbf{F}_{1p}$ and $\mathbf{F}_{2p}$ are given by:

$$\mathbf{F}_{1p} = \begin{bmatrix}\mathbf{H}_k\mathbf{B}_k & \mathbf{0}\\ \mathbf{H}_k(\mathbf{A}_k\mathbf{B}_k + \mathbf{B}_k) & \mathbf{H}_k\mathbf{B}_k\end{bmatrix}, \qquad \mathbf{G}_p = \begin{bmatrix}\mathbf{H}_k\mathbf{A}_k\\ \mathbf{H}_k\mathbf{A}_k^2\end{bmatrix}, \tag{9}$$

$$\mathbf{F}_{2p} = \begin{bmatrix}\mathbf{H}_k\mathbf{B}_k\\ \mathbf{H}_k(\mathbf{A}_k\mathbf{B}_k + \mathbf{B}_k)\end{bmatrix}. \tag{10}$$

If

$$J = \frac{1}{2}\sum_{j=1}^{N}\Big(y_d(k+j) - \hat{y}(k+j)\Big)^{\mathrm{T}}\mathbf{Q}_p\Big(y_d(k+j)) - \hat{y}(k+j)\Big) + \sum_{j=1}^{N}\Big(\Delta u_{mpc}(k+j)\Big)^{\mathrm{T}}\mathbf{R}_p(\Delta u_{mpc}(k+j)), \tag{11}$$

where $y_d(k+j)$, $j = 1, 2, \ldots, N$ is the pressure reference profile, N is the prediction horizon, and $\mathbf{Q}_p$ and $\mathbf{R}_p$ are non-negative definite matrices. Index (17) can be written as

$$J = \frac{1}{2}\hat{\mathbf{Y}}_h^T(k)\mathbf{Q}_p\hat{\mathbf{Y}}_h(k) + \frac{1}{2}\Delta\mathbf{U}_{mpc}^T(k)\mathbf{R}_p\Delta\mathbf{U}_{mpc}(k), \tag{12}$$

where $\hat{\mathbf{Y}}_h^T(k)$ represents the error between the desired output and the predicted output. The solution minimising performance index (12) may then be obtained by solving

$$\frac{\partial J}{\partial\Delta\mathbf{U}_{mpc}} = 0. \tag{13}$$

An off-line computation of the solution may be obtained in an explicit form as follows:

$$\Delta\mathbf{U}_{mpc} = (\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{F}_{1p} + \mathbf{R}_p)^{-1}\Big(\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p(\mathbf{Y}_{d_p}(k) - \mathbf{G}_p\mathbf{z}(k) - \mathbf{F}_{2p}u_{mpc}(k-1))\Big), \tag{14}$$

where $\mathbf{Y}_{d_p}(k)$ is the desired output column vector. For further details see Sunan et al. [10].

4 A stability sufficient constructive condition in GMPC

Theorem 1 *Let us consider the discrete SISO linear system:*

$$\mathbf{z}(k+1) = \mathbf{A}_k\mathbf{z}(k) + \mathbf{B}_k u_{mpc}(k), \tag{15}$$

$$y(k) = \mathbf{H}_k\mathbf{z}(k), \tag{16}$$

obtained by a discretisation of a linear continuous system by means of a sampling time equal to T_s. $u_{mpc}(k)$ is the first element of the

vector of the optimal solution [10] for the GMPC considering the cost function:

$$J = \frac{1}{2}\sum_{j=1}^{N}\Big(y_d(k+j)-\hat{y}(k+j)\Big)^{\mathrm{T}}\mathbf{Q}_p\Big(y_d(k+j)-\hat{y}(k+j)\Big)+ \sum_{j=1}^{N}\Big(\Delta u_{mpc}(k+j-1)\Big)^{\mathrm{T}}\cdot\mathbf{R}_p\Delta u_{mpc}(k+j-1), \quad (17)$$

where $y_d(k+j)$, $j=1,2,\ldots,N$ is the position reference trajectory, N is the prediction horizon and $\mathbf{Q}_p$ and $\mathbf{R}_p$ are non-negative definite matrices. The solution minimising the performance index (17) may be obtained by solving

$$\frac{\partial J}{\partial \Delta \mathbf{U}_{mpc}} = 0. \quad (18)$$

From Sunan et al. [10], it can be seen that the optimal solution is:

$$u_{mpc}(k) = (\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{F}_{1p}+\mathbf{R}_p)^{-1}\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p \Big(\mathbf{Y}_{d_p}(k)-\mathbf{G}_p\mathbf{z}(k)-\mathbf{F}_{2p}u_{mpc}(k-1)\Big), \quad (19)$$

where $\mathbf{Y}_{d_p}(k)$ and $\mathbf{Y}_p(k)$ are the desired output column vector and the measured or observed output vector. Matrices $\mathbf{Q}_p$ and $\mathbf{R}_p$ are diagonal and positively defined. Under the technical hypotheses that $\mathbf{Q}_p=\mathbf{I}$ and $\mathbf{H}_k^{\mathrm{T}}\mathbf{H}=\mathbf{I}$, and the assumption

i) $r_{(1,1)} >> T_s^2$, where $r_{(1,1)}$ represents the first diagonal element of matrix $\mathbf{R}_p$,

then $\forall\ r_{(1,1)}$ such that:

$$r_{(1,1)} > \mathbf{B}_k\mathbf{B}_k^{\mathrm{T}}\frac{\|\mathbf{A}_k\|_2}{1-\|\mathbf{A}_k\|_2}, \quad (20)$$

where $\|\mathbf{A}_k\|_2$ represents the maximal eigenvalue of matrix $\sqrt{\mathbf{A}^T\mathbf{A}}$ and

$$\|\mathbf{A}_k-\mathbf{B}_k\Big((\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{F}_{1p}+\mathbf{R}_p)^{-1}\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{G}_p\Big)\|_2 < \|\mathbf{A}_k+\mathbf{B}_k\Big((\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{F}_{1p}+\mathbf{R}_p)^{-1}\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{G}_p\Big)\|_2, \quad (21)$$

then the system (15) is asymptotically stable.

Proof Theorem 1 *For the sake of brevity, just one prediction step is considered, then:*

$$\mathbf{F}_{1p} = \Big[\mathbf{H}_k\mathbf{B}_k\Big], \quad (22)$$

$$\mathbf{F}_{2p} = \Big[\mathbf{H}_k\mathbf{B}_k\Big], \quad (23)$$

$$\mathbf{G}_p = \Big[\mathbf{H}_k\mathbf{A}_k\Big]. \quad (24)$$

The combination of Eq. (15) with (19) gives this expression:

$$\mathbf{z}(k+1) = \mathbf{A}_k\mathbf{z}(k)+\mathbf{B}_k\Big((\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{F}_{1p}+\mathbf{R}_p)^{-1}\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p \Big(\mathbf{Y}_{d_p}(k)-\mathbf{G}_p\mathbf{z}(k)-\mathbf{F}_{2p}u_{mpc}(k-1)\Big)\Big), \quad (25)$$

which can be written:

$$\mathbf{z}(k+1) = \Big(\mathbf{A}_k-\mathbf{B}_k\Big((\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{F}_{1p}+\mathbf{R}_p)^{-1}\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{G}_p\Big)\Big)\mathbf{z}(k)+ \mathbf{B}_k\Big((\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{F}_{1p}+\mathbf{R}_p)^{-1}\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\Big(\mathbf{Y}_{d_p}(k)\Big)\Big) - \mathbf{B}_k\Big((\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{F}_{1p}+\mathbf{R}_p)^{-1}\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\Big(\mathbf{F}_{2p}(k)\Big)\Big). \quad (26)$$

If

$$r_{(1,1)} > \mathbf{B}_k\mathbf{B}_k^{\mathrm{T}}\frac{\|\mathbf{A}_k\|_2}{1-\|\mathbf{A}_k\|_2}, \quad (27)$$

considering that scalar $r_{(1,1)}>0$ and scalar $\mathbf{B}_k\mathbf{B}_k^{\mathrm{T}}>0$:

$$0 < \|\mathbf{A}_k\|_2 + r_{(1,1)}^{-1}\mathbf{B}_k\mathbf{B}_k^{\mathrm{T}}\|\mathbf{A}_k\|_2 < 1. \quad (28)$$

Recalling that $\mathbf{H}_k^{\mathrm{T}}\mathbf{H}=\mathbf{I}$

$$0 < \|\mathbf{A}_k\|_2 + r_{(1,1)}^{-1}\mathbf{B}_k\mathbf{B}_k^{\mathrm{T}}\mathbf{H}_k^{\mathrm{T}}\mathbf{H}\|\mathbf{A}_k\|_2 < 1 \quad (29)$$

and thus

$$0 < \|\mathbf{A}_k\|_2 + r_{(1,1)}^{-1}\|\mathbf{B}_k\mathbf{B}_k^{\mathrm{T}}\mathbf{H}_k^{\mathrm{T}}\mathbf{H}\mathbf{A}_k\|_2 < 1. \quad (30)$$

With matrix $\mathbf{F}_{1p}$ defined as in (22) and $\mathbf{G}_{1p}$ defined as in (24), it is known that matrix $\mathbf{B}_k$ is proportional to T_s, then considering that $\mathbf{R}_p = r_{(1,1)}$, choosing a suitable $r_{(1,1)} >> T_s$ and considering that $\mathbf{Q}_p=\mathbf{I}$ (technical hypothesis), this condition is derived:

$$0 < \|\mathbf{A}_k\|_2 + \Big\|\Big(\mathbf{B}_k\Big((\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{F}_{1p}+\mathbf{R}_p)^{-1}\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{G}_p\Big)\Big\|_2 < 1. \quad (31)$$

Considering the norm properties and condition (21), then:

$$\begin{aligned} 0 &< \Big\|\mathbf{A}_k-\mathbf{B}_k\Big((\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{F}_{1p}+\mathbf{R}_p)^{-1}\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{G}_p\Big)\Big\|_2 \\ &< \Big\|\mathbf{A}_k+\mathbf{B}_k\Big((\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{F}_{1p}+\mathbf{R}_p)^{-1}\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{G}_p\Big)\Big\|_2 \\ &< \|\mathbf{A}_k\|_2 + \Big\|\mathbf{B}_k\Big((\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{F}_{1p}+\mathbf{R}_p)^{-1}\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{G}_p\Big)\Big\|_2 < 1. \end{aligned} \quad (32)$$

To conclude

$$0 < \|\mathbf{A}_k\| - \Big\|\mathbf{B}_k\Big((\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{F}_{1p}+\mathbf{R}_p)^{-1}\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{G}_p\Big)\Big\|_2 < 1. \quad (33)$$

□

The constraint in (27) states a plausible condition on the controller.

If mass $m \to \infty$, then, because of the discretisation and according to the Landau notation, $\Big(\Big\|\lambda_{max}\Big(\mathbf{A}_k\Big)-1\Big|\Big) \to 0$ with $O\Big(\Big|\lambda_{max}\Big(\mathbf{A}_k\Big)-1\Big|\Big) = O(\frac{1}{m})$. In the meantime, $O(\mathbf{B}_k\mathbf{B}_k^{\mathrm{T}}) = O(\frac{1}{m^2})$. For a very slow system, according to (27), $r_{(1,1)} \to 0$, parameter $r_{(1,1)}$ is present in the denominator function of the optimal solution in (19) and small values of $r_{(1,1)}$ are devoted to speeding up the system. If mass $m \to 0$, then, because of the discretisation, $O\Big(\Big|\lambda_{max}\Big(\mathbf{A}_k\Big)\Big|\Big) = O\Big(\frac{1}{m}\Big)$, but in the meantime $\mathbf{B}_k\mathbf{B}_k^{\mathrm{T}} \to \infty$ with $O(\mathbf{B}_k\mathbf{B}_k^{\mathrm{T}}) = O(\frac{1}{m^2})$. For a very fast system, $r_{(1,1)} \to \infty$, parameter $r_{(1,1)}$ is devoted to slowing down the system. We can therefore conclude that a highly inertial system needs relatively small values of $r_{(1,1)}$ to be optimised and stabilised. If the inertia is small, then the system needs larger values of $r_{(1,1)}$ to be optimised and stabilised. In fact, very fast systems can have very high abrupt changes in the input signals and in the cost function, and the input factor needs to be reduced to find an optimality.

5 The case of input saturation

Proposition 1 *If the discrete SISO linear system is considered:*

$$\mathbf{z}(k+1) = \mathbf{A}_k\mathbf{z}(k) + \mathbf{B}_k u_{mpc}(k), \quad (34)$$

$$y(k) = \mathbf{H}_k\mathbf{z}(k), \quad (35)$$

and

$$|u_{mpc}(k)| \leq U_{max} \ \ \forall k, \quad (36)$$

then (34) the input saturation defined in (36) is asymptotically stable and its input avoiding the constraint if condition (27) holds together with the input limitation. The following condition summarises the result:

$$r_{(1,1)} > \max\left\{\frac{\mathbf{B}_k\mathbf{B}_k^{\mathrm{T}}\|\mathbf{Y}_{d_p}(k)\|_2}{U_{max}} + \frac{\left\|\mathbf{B}_k\mathbf{F}_{1p}\mathbf{Q}_p\mathbf{F}_{2p}\right\|_2}{U_{max}}|u_{mpc}(k-2)|, \ \ \mathbf{B}_k\mathbf{B}_k^{\mathrm{T}}\frac{\|\mathbf{A}_k\|_2}{1-\|\mathbf{A}_k\|_2}\right\}. \quad (37)$$

Proof Proposition 1 *The demonstration is straightforward just considering that*

$$|u_{mpc}(k)| < |u_{mpc}(k)| < U_{max} \ \ \forall k$$

and thus it is enough that the following condition holds:

$$\left\|\mathbf{B}_k\Big((\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{F}_{1p} + \mathbf{R}_p)^{-1}\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\Big(\mathbf{Y}_{d_p}(k)\Big)\Big) - \mathbf{B}_k(\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\mathbf{F}_{1p} + \mathbf{R}_p)^{-1}\mathbf{F}_{1p}\mathbf{Q}_p\mathbf{F}_{2p}u_{mpc}(k-1)\right\|_2 \leq U_{max}. \quad (38)$$

In fact, using similar considerations as before, the following expression is obtained:

$$\left\|\mathbf{B}_k r_{11}^{-1}\mathbf{F}_{1p}^{\mathrm{T}}\mathbf{Q}_p\Big(\mathbf{Y}_{d_p}(k)\Big)\right\|_2 + |\mathbf{B}_k\mathbf{F}_{1p}\mathbf{Q}_p\mathbf{F}_{2p}| \leq U_{max}, \quad (39)$$

and thus, including also the stability condition, condition (37) follows here again:

$$r_{(1,1)} > \max\left\{\frac{\mathbf{B}_k\mathbf{B}_k^{\mathrm{T}}\|\mathbf{Y}_{d_p}(k)\|_2}{U_{max}} + \frac{\left\|\mathbf{B}_k\mathbf{F}_{1p}\mathbf{Q}_p\mathbf{F}_{2p}\right\|_2}{U_{max}}|u_{mpc}(k-1)|, \ \ \mathbf{B}_k\mathbf{B}_k^{\mathrm{T}}\frac{\|\mathbf{A}_k\|_2}{1-\|\mathbf{A}_k\|_2}\right\}. \quad (40)$$

□

It is possible to observe that for large values of U_{max}, the condition on the input barrier in the cost function (17) with weight $r_{(1,1)}$ is not so restrictive, so larger inputs are allowed. For small values of U_{max}, the input barrier limits the values of the input and no large input values are allowed.

6 Simulation results

It must be clarified that function $m_{in}(t)$ is a stepwise constant function with $m_{in}(t) = 0.086$ (kg/sec) or $m_{in}(t) = 0$ and in the simulated case $m_{in}(t) = 0.086$ (kg/sec) is considered. Two cases should be differentiated: weak anti-saturating action:

$$r_{(1,1)} > \frac{\mathbf{B}_k\mathbf{B}_k^{\mathrm{T}}\|\mathbf{Y}_{d_p}(k)\|_2}{U_{max}}, \quad (41)$$

and strong anti-saturating action:

$$r_{(1,1)} >> \frac{\mathbf{B}_k\mathbf{B}_k^{\mathrm{T}}\|\mathbf{Y}_{d_p}(k)\|_2}{U_{max}}. \quad (42)$$

Figure 2 shows the controlled pressure, which represents the result. If the anti-windup action is relatively weak, more time is necessary to re-establish the control loop. As already explained, during the windup effect, the feedback control is broken. The long period where the pressure is at negative values can be explained by the absence of feedback control action. From Fig. 3, representing the mass flow $\frac{dm(t)}{dt} = m_{in}(t) - m_0(t)$ (kg/sec), it can be seen that this function is a consequence of the relation:

$$\frac{dm(t)}{dt} \approx \frac{dp(t)}{dt}. \quad (43)$$

After taking the following equation into consideration one more time:

$$\frac{dm(t)}{dt} = m_{in}(t) - m_0(t), \quad (44)$$

for activating the process, a strong initial action through the mass flow $m_0(t)$ is necessary. These two figures show that in case of a strong anti-saturating action in Fig. 5, the controlled system emerges from the saturation state very quickly with faster dynamics

thanks to the stronger anti-windup action. After saturation occurs, the control loop becomes open and there is no presence of feedback control.

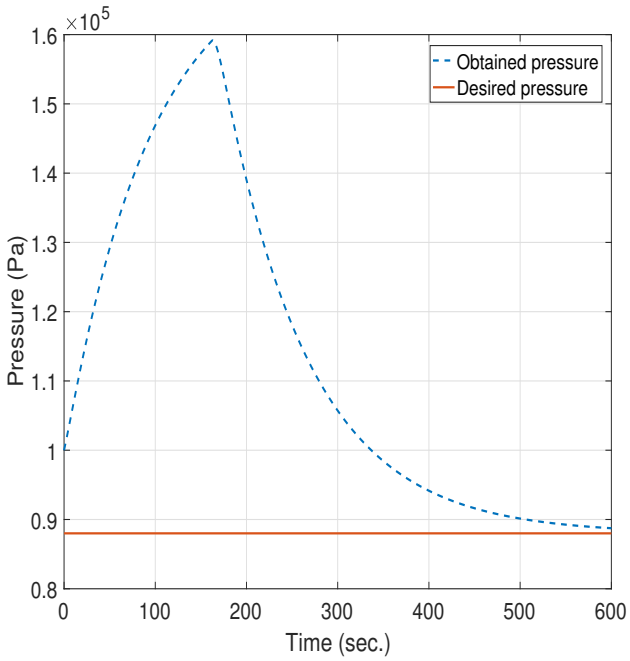


Figure 2: Desired and obtained pressure with weak anti-saturating action

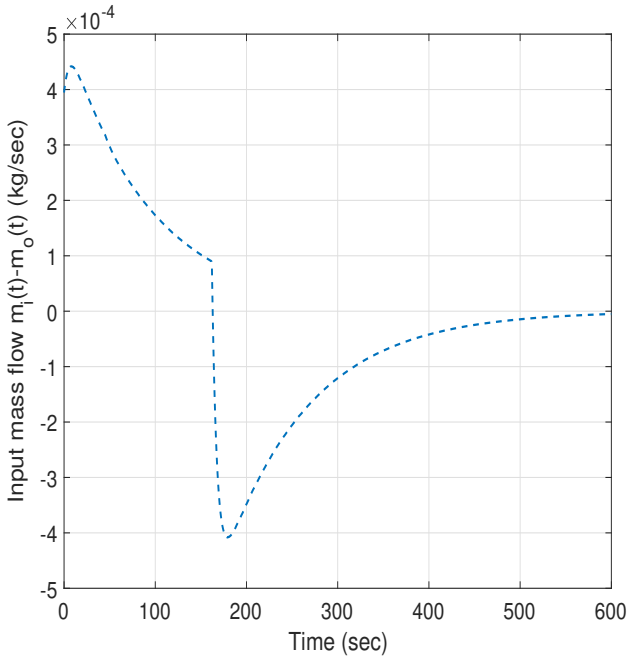


Figure 3: Mass flow $\frac{m(t)}{dt} = m_{in}(t) - m_0(t)$ (kg/sec) with weak anti-saturating action

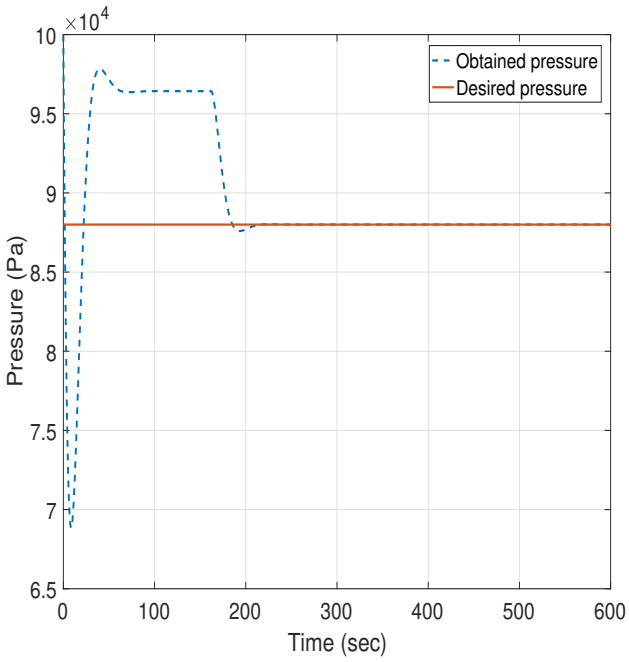


Figure 4: Desired and obtained pressure with strong anti-saturating action

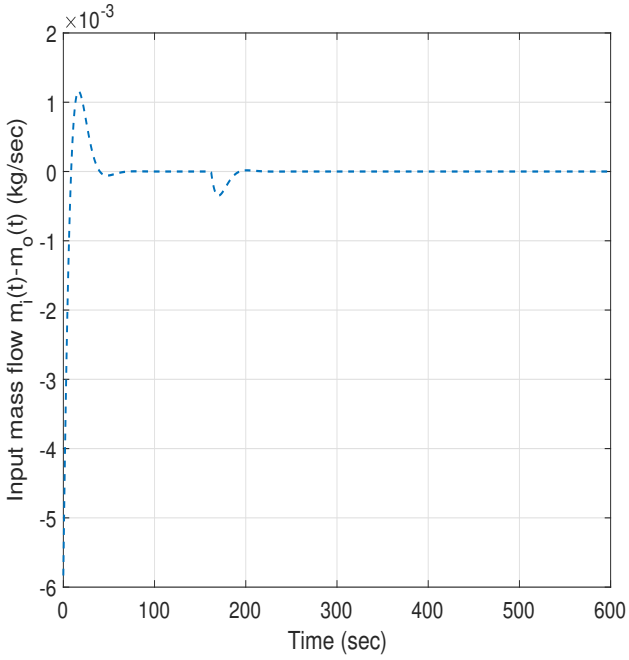


Figure 5: Mass flow $\frac{dm(t)}{dt} = m_{in}(t) - m_0(t)$ (kg/sec) with strong anti-saturating action

7 Conclusion

Conservative conditions for stability are a crucial problem in optimisation using LMPC. This contribution is devoted to a sufficient and constructive condition for the stability of an LGMPC, which calculates a lower bound for the elements of matrix **R**. The obtained results are physically interpreted. An illustrative example is provided in which a water recovery process is taken into consideration to test the proposed results through computer simulations.

Conflict of Interest The authors declare no conflict of interest.

References

[1] Bolognani, S.; Peretti, L.; Zigliotto, M. Design and implementation of model predictive control for electrical motor drives. *IEEE Transactions on Industrial Electronics* **2009**, *56*, pp. 1925–1936.

[2] Neelakantan, V.; Washington, G.; Bucknor, N. Model Predictive Control of a Two Stage Actuation System using Piezoelectric Actuators for Controllable Industrial and Automotive Brakes and Clutches. *J. Intell. Mater. Syst. and Struct.* **2008**, *19*, pp. 845–857.

[3] Hu, Z.; Farson, D. Design of a waveform tracking system for a piezoelectric actuator. Proc. Inst. Mech. Eng., J. Syst. Control Eng., 2008, Vol. 222, pp. 11–21.

[4] Bemporad, A.; Morari, M.; Dua, V.; Pistikopoulos, E.N. The explicit linear quadratic regulator for constrained systems. *Automatica* **2002**, *38*, pp. 3–20.

[5] Pedret, C.; Poncet, A.; Stadler, K.; Toller, A.; Glattfelder, A.; Bemporad, A.; Morari, M. Model-varying predictive control of a nonlinear system. *Internal report in Computer Science Dept. ETSE de la Universitat Autònoma de Barcelona* **2000**.

[6] Liu, Y.C.; Lin, C.Y. Model Predictive Control with Integral Control and Constraint Handling for Mechatronic Systems. Proc. of the 2010 International Conference on Modelling, Identification and Control; , 2010; pp. 424–429.

[7] Mercorelli, P.; Goes, J.; Halbe, R. A Lyapunov based PI controller with an anti-windup scheme for a purification process of potable water. Proc. of the 2014 International Conference on Control, Decision and Information Technologies (CoDIT), 2014, pp. 578–583.

[8] Mercorelli, P. An Optimal and Stabilising PI Controller with an Anti-windup Scheme for a Purification Process of Potable Water. Proc. of the 16th IFAC Workshop on Control Applications of Optimization CAO'2015, 2015, pp. 259–264.

[9] Mercorelli, P. A sufficient asymptotic stability condition in generalised model predictive control to avoid input saturation. *Lecture Notes in Electrical Engineering* **2019**, *489*, pp. 251–257.

[10] Sunan, H.; Kiong, T.; Heng, L. *Applied Predictive Control*; Springer-Verlag London: Printed in Great Britain, 2002.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 321-326 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Estimating Academic results from Trainees' Activities in Programming Exercises Using Four Types of Machine Learning

Toshiyasu Kato*[1], Yuki Terawaki[2], Yasushi Kodama[2], Teruhiko Unoki[3], Yasushi Kambayashi[1]

[1]*Nippon Institute of Technology, Saitama, Japan*

[2]*Hosei University, Tokyo, Japan*

[3]*Photron Limited, Tokyo, Japan*

ARTICLE INFO

Article history:
Received: 28 May, 2019
Accepted: 09 October, 2019
Online: 22 October, 2019

Keywords:
Programming Exercises
Classroom Assistants
Learning analytics
Machine Learning
Grading estimate

ABSTRACT

Predicting trainees' final academic results in the early stage of programming class is a significant mission in the field of learning analytics. Performing exercises in programming class is hard and it takes a lot of time for trainees. For this reason, careful support with trainees are offered in many classes through classroom assistants (CAs). Even with CAs' assistances, managing a programming class is difficult. Because each trainee's coding activity is different from another's, even when each of them is solving the same exercise. Classroom assistants with little teaching experience have difficulty for understanding the unique features of trainee's coding activity. We have employed data mining to analyze trainees' coding activities to distinguish those various features. The objective of this research is to present such behavioral features of trainees to CAs to enrich their assistance for the trainees. In order to investigate the timing of guidance, we conjectured the academic results from the chronicle record of coding activities.

1. Introduction

This paper is an extension of the work originally presented in 2017 International Conference on Computational Science and Computational Intelligence (CSCI) [1].

In programming exercises, Individual trainee's coding activity varies from person to person, even when each trainee is answering the common exercise. The aim of this research is to display many coding features of trainees to classroom assistants (CAs). We conjectured the academic results to indicate the timing of instruction to the CAs.

Through the analysis of the logs of trainees by using the data mining techniques, we can find the learning pattern of each trainee. This analysis helps instructors to identify "at-risk" trainees [2]. Programming exercises in higher education institutions allow trainees to solve individual tasks. Since the skill levels of trainees differ, each of them has to work on tasks in an individual way [3]. Some trainees solve exercises easily, while others take much time. Therefore, many schools employ CAs to provide assistances for the trainees. CAs are expected not only answer questions of trainees but also to provide comprehensive assistances for the trainees with large assignments [4]. For this reason, a CA needs to understand who needs assistance, what kind of assistance the trainee needs, and under what circumstances assistance the trainee confusing [5].

Predicting trainees' final academic results in the early stage of programming class is a *significant mission* in the field of learning analytics. For instance, one can perform regression analysis for this purpose [6]. In order to catch the timing when the instructors intervene to the trainees who need help, instructors need to conjecture the trainees' future scores of the programming exercises. In order to provide such estimations, the authors have developed and used a unique programming exercise support system [7]. The system is called PRESS, and through the analysis of the outputs of PRESS, we have succeeded to identify the coding activities.

Trainees' progresses differ from each other, because of the difference of individual skill levels [3]. We have found that this skill level difference can be observed in the coding mode. Trainees who show slow progress have poor coding modes. PRESS demonstrates that such situations are common. Such trainees often

*Toshiyasu Kato, Nippon Institute of Technology 4-1 Gakuendai, Miyashiro-machi Minami-saitama-gun Saitama, 345-8501, Japan, katoto@nit.ac.jp

stop programming and become disgusted [8]. Therefore, it is important for instructors to identify what kind of trainees have programming exercises at an early stage [4].

We have conjectured the trainee's academic results in order to get the timing when CAs start assistance for the problematic trainees. According to the instructor's opinion, trainees who take a lot of time for exercises tend to have low academic results. In this paper, we give details of the results of evaluation experiments on this estimation. As a result, CAs became to comprehend the trainee's coding activities without having prior programming support experience.

2. Related Work

Several researchers have reported that coding activities has improved trainee's skills by assisting in debugging and coding methods. Ryan has built a model that enables trainees to improve their skills by analyzing debug logs and execution results [9]. Our study analyzes coding history in addition to debug logs.

Alex proposed a tutor on a computer to support trainee programming [10]. The tutor checks whether the trainees are on the correct trajectory and advises trainees in a wrong track. For instance, an instructor may give trainees a hint on how to reconstruct their programs. Our research aim is to improve trainees' practical skills by supporting coding methods.

There are past researches targeted at study how to support trainees through their learning chronicle data. In such studies, scientists have analyzed teaching materials using techniques based on statistical methods on Support Vector Machines (SVM). For instance, Yamada et al. detected contents on the Web pages that were difficult for trainees to comprehend by logistic regression [11]. In that research, they have created models with seven explanatory parameters, namely latent time, gaze point distance moved, gazing point movement speed, mouse distance moved, mouse speed, the amount of mouse wheel rotation, and correct identification. Yamada's evaluation experiment have achieved the identification rate of 81% by using SVM and time series data to improve coding activities, such as compilation and execution frequency.

Researches on foreseeing trainee's performance by machine learning (ML) algorithm that takes advantages of trainee's learning chronicle data as teaching data are as follows.

Guo, Zhang, Xu, Shi, and Yang applied deep neural network (DNN) to the data from 12,000 trainees in Chinese elementary and junior high schools to predict five academic results in high school entrance exams [12]. They found that deep learning can predict academic results with highest accuracy rate (Accuracy) among many ML methods.

Belachew and Gobena employed three types of ML algorithms, namely neural networks, Naive Bayes and support vector machines to predict results of 993 trainees at Wolkite University in Ethiopia [13]. They found that, as a result of comparison, Naive Bayes had the highest accuracy rate.

Ahmad, Ismail, and Aziz used three ML algorithms, namely Naive Bayes, decision trees, and rule bases to predict the performance of about 500 trainees in Malaysian [14]. As a result, they reported that the rule base has the highest accuracy rate.

Amrieh, Hamtini, and Aljarah used three ML algorithms, namely neural nets, decision trees, and Naive Bayes, to predict performance with data from 480 trainees at Jordanian University [15]. They found that decision trees and neural networks obtained the highest accuracy rates.

In these preceding researches, variety of ML algorithms are applied to trainee data, and researchers who tackled this problem demonstrated how and what way each of the ML methods succeeds to predict the final scores. Unfortunately, the data used for the evaluation are largely different from each other, and the data used in many studies are not open to the public. It is difficult for a third party to pursue further study through conducting a comparative experiments with a follow-up test or alternative method. We can only use them as reference data.

3. Analysis of Coding Activities

In this research, we have set up data on the coding activities of trainees in 2014 and 2015 classes analyzed in previous study [16, 17]. Table 1 displays the explanatory parameters used for the analysis. We have measured all of the data and then analyzed the analysis targets using PRESS. We have examined the following two approaches to the data.

- Unsupervised classification.
- Estimation of exercise-solving time.

Table 1: Explanatory parameters

Type of characteristics		
Time	Frequency	Error
Problem-Solving	Compilation	Number of errors
Compilation interval	Execution	Number of similar errors
Execution interval		

3.1. Unsupervised classification

We have used Ward's method, which is a representative hierarchical clustering [18] [19]. Clustering is a process that groups objects into classes of similar objects [20]. We have formed from

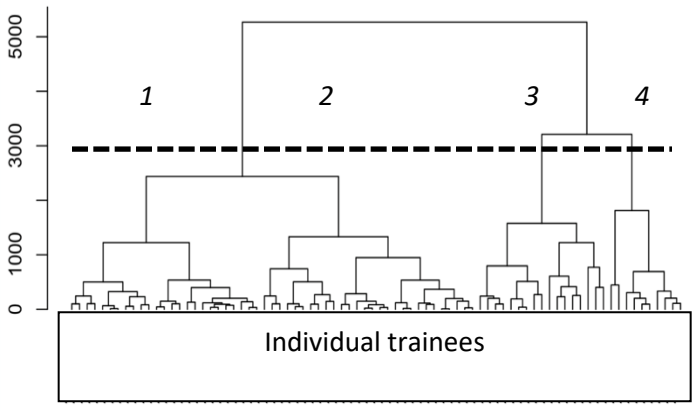


Figure 1: Clustering by Ward's method

Table 2: Explanatory parameters for each cluster

Cluster	Solving time [s]	Compilation interval [s]	Execution interval [s]	Compilation frequency	Execution frequency	Number of errors	Number of similar errors	Number of trainees
1	528	209	267	7	5	1	1	25
2	812	379	606	6	3	3	3	28
3	1405	590	1160	7	3	4	3	17
4	1506	228	337	18	13	5	4	10

the patterns (observations, data elements, or characteristic vectors) into clusters (groups or subsets) based on those localities and connectivity in the n-dimensional space as shown in Figure 1. Figure 1 demonstrates the cluster analysis of the seven explanatory parameters. Table 1 indicates the summary of explanatory parameters for each cluster in Figure 1.

We have perceived the following clusters:

- Cluster 1: A group solved the exercise in short time. In this group, trainees understand each problem they encounter while coding.
- Cluster 2: A group solved the exercise in short time. In this group, trainees are coding without understanding the details of their problems.
- Cluster 3: A group solved the exercise in long time. In this group, trainees are cursorily checking the error contents and cording without full comprehension of their problems.
- Cluster 4: A group solved the exercise in long time. In this group, trainees are cording without checking any error details.

3.2. Length estimation of time for solving exercises

The trainees in each cluster should receive cluster-specific instructions. Therefore, instructors and CAs need to identify the features of the trainees in the class and pay special attentions to those who take a long time to solve given exercises. The cluster analysis, however, cannot conjecture the duration for solving exercises in the future. For this reason, CAs cannot know which trainees need certain types of guidance at the beginning or before class. In order to ameliorate this hindrance, we used the 2014 class coding activities as training data and tried to predict the exercise-solving time in 2015 class activities. Table 3 displays the coding activity data for two years. The values of the explanatory parameters for each year are averaged values (excluding the number of trainees). These data are taken from two programming exercises of the same content each year.

We have used logistic regression to decide whether trainees' exercise-solving time would be long or short. Logistic regression is a typical statistical method that illustrates the relationship between two different series of data [21]. We can assume that one of the series of data as independent variables, and the other data are dependent variables. We assume that the output variable takes one value from a set of fixed values. This value corresponds to a class of the classification problem. The regression equation is displayed in Equation (1). p is the probability conjecture of the

$$p = \frac{1}{1+e^{-(\beta_0+\beta_1\chi_1+\beta_2\chi_2+\cdots+\beta_k\chi_k)}} \quad (1)$$

Table 3: Coding activities in 2014 and 2015

	2014	2015
Number of trainees	85	80
Solving time (Problem) [s]	815	936
Compilation interval (Average) [s]	295	352
Execution interval (Average) [s]	513	584
Compile frequency	10	8
Execution frequency	5	5
Number of errors until correct answers	7	3
Number of similar errors that are most frequent	4	3
Number of trainees late submission	32	35
Number of exercise resolutions	40	41

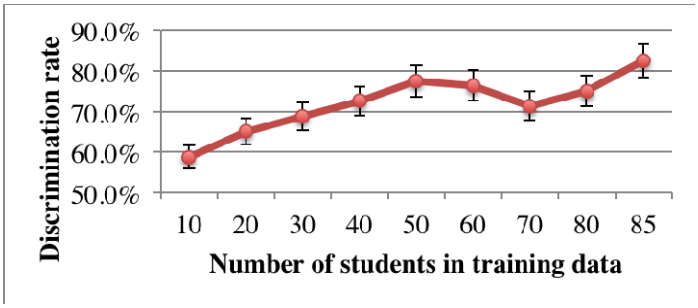


Figure 2: The number of trainees in the training data and discrimination rate

objective variable. In this research, we obtained the probability p of a long exercise-solving time. β_0 is a constant, β_1, β_2, ... β_k are regression coefficients, and $\chi_1, \chi_2, \ldots \chi_k$ are explanatory parameters.

We first applied training data to logistic regression to determine constant and regression coefficients. We used Tensor Flow for this calculation. Tensor Flow is a representative tool for DNN [22]. We used cross-entropy as a learning error function for the supervised classification problem and carried out optimization using the steepest descent method [23]. The objective variable for logistic regression must be represented in the 0/1 binary. We set the teaching data to 0 when it was longer than the mean exercise-solving time and to 1 if it was shorter than the mean exercise-solving time.

We created a model with the 2014 data, and then we applied the 2015 data to the model and conjectured the length of the exercise-solving time. The discrimination rate was 82.5%. The

parameters at this time were a learning rate of 0.015 and the number of learning iterations was 70,000. Figure 2 displays the number of trainees and the identification rate used in the training data for ML.

4. Estimating Academic results from Trainees' Activities

Now that we have the conjecture of the exercise-solving time, we next conjectured academic results to find the timing of instruction. In order to improve accuracy much further, we have used 2016 data in addition to the last two years (Table 3). Table 4 displays the coding activities for 2016 class and Table 5 displays the academic results by three years. The academic results are based on the final exam at the end of the semester.

The subjects (trainees) are sophomores in a Java programming class from our institute. Each year, the same instructors took the charge of the class, but the academic results are differed from year to year as a subjective evaluation of instructors has been added. For this reason, we have defined the estimation rate as follows along with the coincidence rate:

$$Estimation\ rate\ = \frac{Estimated\ academic\ results\ \pm 1}{Number\ of\ trainees}$$

Based on the data of these three years, we have conjectured the academic results through the following procedure.

1. Normalize the input data.
2. Decide the features.
3. Decide network configuration and parameters.
4. Conjecture the accuracy by creating a network model.
5. Check the estimation accuracy.
6. Repeat steps 3 to 5 to improve accuracy of the estimation.
7. Analyze by changing the hidden layer

Table 4: Coding activities in 2016

Number of trainees	107
Solving time (Exercise) [s]	639
Compile interval (Average) [s]	405
Execution interval (Average) [s]	139
Compile frequency	7
Execution frequency	2
Number of errors until correct answers	5
Number of similar errors that are most frequent	2
Number of trainees late submission	51
Number of exercise resolutions	38

Table 5: Academic results in programming exercises

Academic results	2014	2015	2016
1: Failing	17	1	5
2: Below pass	33	13	34
3: Pass	17	17	34
4: Very good	21	36	30
5: Excellent	2	15	7

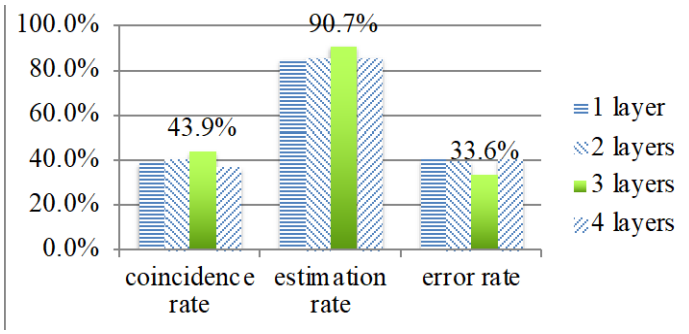


Figure 3: The coincidence rate, estimation rate, and error rate

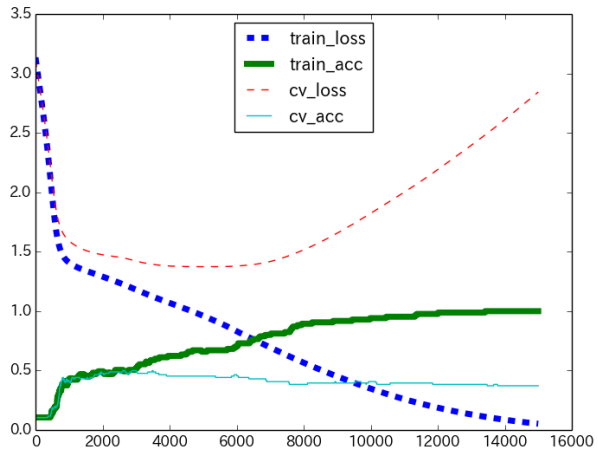


Figure 4: Loss and accuracy when the hidden layer is three layers

Figure 3 displays the coincidence rate, estimation rate, and error rate in the case of one, two, three and four hidden layers respectively. The error rate in Figure 3 is calculated by sum of the divergences between conjectured academic results and final academic results then divided by the number of trainees.

In addition to these parameters, we have chosen the following features to analyze the trainees' activities. They are duration for solving the exercises compile frequency, execution frequency, the number of similar errors that are most frequent, and the number of exercises they resolved. Through various machine-learning situations, we have found that the case of three hidden layers provides the most accurate estimation. Figure 4 displays the best case of our experiments; that has three hidden layers.

Table 6 displays the mean values of explanatory parameters. We have normalized them for each academic result. There was no trainee conjectured to earn the academic result of 5 (highest). In academic result, the number of exercises solved and the time for solving exercises are particularly related.

In addition to the experiments using DNN, we have performed performance prediction using three ML algorithms. The algorithms used in this experiment are Random Forest, SVM, and Naive Bayes. The objective of these experiments is to assess the performance of different ML algorithms.

With those ML algorithms, the purpose is also classification; the same as DNN. Also, we used the same data set for analysis as

Table 6: The mean values of feature quantities normalized for each academic result

Ac. result	Exercise-solving times	Compile times	Execution times	Similar Errors	Number of exercise resolutions
1	1.66	-0.40	-0.96	-0.78	-2.28
2	0.25	0.44	-0.44	0.47	-0.65
3	0.20	-0.77	-0.69	-0.68	-0.61
4	-0.96	-0.32	0.32	-0.30	1.01
5	-	-	-	-	-

Table 7: Estimation results

	Accuracy	Precision	Recall
DNN	0.70	0.41	0.46
Random Forest	0.62	0.34	0.37
SVM	0.58	0.32	0.42
Naive Bayesian	0.58	0.32	0.38

DNN. The reason for adopting these three types is that they were used for performance prediction in the previous researches [11] [13] [14] [15].

Table 7 displays the results of various data mining methods. The interpretation of precision and recall values here depends on each class and each academic result. This value is based on the average of all academic result classes. From these results, we can confirm that the academic results can be conjectured through analysis of coding activities by any MLs. DNN provides the most accurate predictions, but precision and recall values are low. From the perspective of accuracy, our method can find trainees who would fail from the class. However, since precision and recall are low, there is an exercise that the same method does not guarantee the same accuracy every year.

5. Discussion

We have conducted experiments of DNN with various combinations of explanatory parameters. We could not identify one single particular dominating explanatory parameter. We can conclude that all the explanatory parameters contribute almost equally to the accurate estimations. What we can state is that the DNN provides the highest values as displayed in Table 7.

For this data set, we can conclude that we can predict the performance of each trainee by any algorithms we select, even though the Deep Leaning provides the best results. However, precision and recall values are low even in DNN. The reason may be that the trainees are not the same every year and the final exam is different each year. The low value of precision and recall can be explained that the past trainee's patterns were not suitable for the target trainee's patterns. This can be expected to improve accuracy by increasing the number of trainee patterns each year.

Regarding the explanatory parameters, the most relevant activities to the academic results are the frequency of compilations and the number of similar errors. Trainees who display too frequent compilation and produce a number of similar errors are expected to receive low academic results. We can conclude that these trainees are doing programming without knowing what wrong they have done in the programming. However, there were some trainees with low academic results despite the small amount of compilations. There is a possibility that these trainees just copied their programs from other trainees. Instructor should intervene trainees those who compile too frequently in a short period of time. Also, we have observed that intervention is necessary for trainees who have not compiled for a long time. From these observation, we can conjecture that not only we can build a system that let instructors know which trainees need interventions but also we can construct a computerized automatic intervention system for trainees who need help.

We interpret the results in Table 6 and consider each group of trainees has the following features.

- Ac. result 1: A group solved the exercises in short period and produced a small number of executions. Some trainee in this group may be programming carefully, but in practice most of them are less motivated and not programming at all. This group of trainees may be waiting for their friends' solutions of the task.
- Ac. result 2: A group compiled the programs very frequently, causing the similar errors. This group of trainees may not investigate the cause of the error very well, and may just execute their program mindlessly. Because this type of trainees do not understand the cause of the error, the final exam scores also tend to be low.
- Ac. result 3: A group produced a small number of compilations and executions**.** In this group, trainees do not produce the same type of errors many times. Trainees are coding while they are investigating the cause of the error carefully. They are trying to understand every error they produced. Therefore, their final exam scores tend to be high.
- Ac. result 4: A group consumed very short period for solving the exercises. Trainees in this group solve many exercises. We observed trainees in this group can understand exercise sentences quickly and can program for solutions quickly. Therefore, their final exam scores tend to be very high.

The numbers of incorrect estimations are as follows. There were six cases where the conjectured academic result was 4 and the actual academic result was 2. Also, there were four cases where the conjectured academic result was 2 and the actual academic result was 4.

The reason why there were no trainees of academic result 5 in Table 6 may be that the number of exercises solved was small. Actually, we do not have enough training data with trainees with this highest academic result.

6. Conclusion

In this paper, we have analyzed the trainee's coding history to conjecture exercise-solving time and academic results. The objective of this estimation is to support CAs in order to assist trainees in programming exercises. We have carried out the analyses to classify trainees who perform programming exercises.

We have applied ML to construct a model that represents the coding activities of a programming class and then applied that model to the coding activities of trainees in the following year to conjecture the results from the data for the past two years. As the results, we have observed that the coincidence rate is 43.9%, and the plus or minus 1 estimation rate was 90.7%. In addition, in the analysis using four types of ML, we have found that DNN provides the most accurate estimation. However, since the values of precision and recall are low, it is necessary for us to investigate the coding history of the trainees of each year, and then improve our analysis.

As the data mining, we discovered that DNN, Random Forest, SVM and Naive Bayes can foresee the same results with sufficiently high accuracy by the same data. We clarified the results and found that we could conjecture the academic results even with a small amount of explanatory parameters.

As a future direction for research, we analyze each lesson more methodically so that we can conjecture the results at an early stage while maintaining the estimation rate high. We also plan to apply this method to other programming classes that use other programming languages than Java, such as Python.

Acknowledgment

This work was supported by Japan Society for Promotion of Science (JSPS), with the basic research program (C) (18K02910), and (A) 26245048, Grant-in-Aid for Scientific Research (KAKENHI).

References

[1] Kato, T., Kambayashi, Y., Terawaki, Y. & Kodama, Y., “Estimating Grades from Students' Behaviors in Programming Exercises Using Deep Learning”, In 2017 International Conference on Computational Science and Computational Intelligence (CSCI), IEEE, 1115-1119, 2017.

[2] Baradwaj, B. & Pal, S., “Mining Educational Data to Analyze Student’s Performance”, International Journal of Advanced Computer Science and Applications, 6, 2, 63-69, 2011.

[3] Horiguchi, S., Igaki, H., Inoue, A., Yamada M., Hoshi T. & Okada K., "Progress Management Metrics for Programming Education of HTML-based Learning Material", Journal of Information Processing Society of Japan., 53, 1, 61-71, 2012. (In Japanese)

[4] Sagisaka T. & Watanabe S., "Investigations of Beginners in Programming Course Based on Learning Strategies and Gradual Level Test, and Development of Support-Rules", Journal of Japanese Society for Information and Systems in Education., 26, 1, 5-15, 2009. (In Japanese)

[5] Yasuda K., Inoue A. & Ichimura S., "Programming Education System that can Share Problem-solving Processes between Students and Teaching Assistants", Journal of Information Processing Society of Japan., 53, 1, 81-89, 2012. (In Japanese)

[6] You, J. W., “Identifying significant indicators using LMS data to predict course achievement in online learning”, Internet and Higher Education, 29, 23-30, 2016.

[7] Kato T. & Ishikawa T., "Design and Evaluation of Support Functions of Course Management Systems for Assessing Learning Conditions in Programming Practicums", ICALT 2012, pp. 205-207, 2012.

[8] Igaki. H, Saito S., Inoue A., Nakamura R., & Kusumoto S., "Programming Process Visualization for Supporting Students in Programming Exercise", Journal of Information Processing Society of Japan, 54, 1, 2013. (In Japanese)

[9] Ryan C. & Michael C. L., "Debugging: from Novice to Expert", ACM SIGCSE Bulletin: 36, 1, 17-21, 2004.

[10] Alex G., Johan J. & Bastiaan H., "An Interactive Functional Programming Tutor", ITiCSE '12: Proceedings of the 17th. ACM, 250-255, 2012.

[11] Yamada T., Nakamichi N. & Matsui T., "Detection of Low Usability Web Pages with Logistic Regression", GN Workshop 2010, 57-62, 2010. (In Japanese)

[12] Guo, B., Zhang, R., Xu, G., Shi, C., & Yang, L., "Predicting Students Performance in Educational Data Mining", 2015 International Symposium on Educational Technology (ISET), doi:10.1109/iset.2015.33, 2015.

[13] Belachew, E. B., & Gobena, F. A., "Student Performance Prediction Model using Machine Learning Approach: The Case of Wolkite University", International Journal of Advanced Research in Computer Science and Software Engineering, 7, 2, 46-50. doi:10.23956/ijarcsse/v7i2/01219, 2017.

[14] Ahmad, F., Ismail, N. H., & Aziz, A. A., "The Prediction of Students Academic Performance using Classification Data Mining Techniques", Mathematical Sciences 9, 6415-6426. doi:10.12988/ams.2015.53289, 2015.

[15] Amrieh, E. A., Hamtini, T., & Aljarah, I., "Preprocessing and Analyzing Educational Data Set using X-API for Improving Students Performance", 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), doi:10.1109/aeect.2015.7360581, 2015.

[16] Kato T., Kambayashi Y. & Kodama Y., "Data Mining of Students' Behaviors in Programming Exercises", Smart Education and e-Learning 2016. 59, 121-133, 2016.

[17] Kato T., Kambayashi Y., Terawaki Y. & Kodama Y., " Analysis of Students' Behaviors in Programming Exercises using Deep Learning", Smart Education and e-Learning 2017. 38-47, 2017.

[18] Kamishima T., "A Survey of Recent Clustering Methods for Data Mining (Part 1): Try Clustering!", Journal of the Japanese Society for Artificial Intelligence, 18, 1, 59-65, 2003. (In Japanese)

[19] Michael R. A., "Cluster Analysis for Applications", Academic Press, 1973.

[20] Jain, A. K., Murty, M. N. & Flynn, P. J., "Data Clustering: A review", ACM Computing Surveys. 31, 3, 264-323, 1999.

[21] Uchida O., "Logistic Regression Analysis by SPSS", Ohmsha, 2011. (In Japanese)

[22] Tensor Flow, Google, https://www.tensorflow.org

[23] Ozawa Y., Kohno Y. & Takahashi T., "Generalization method with satisficing on ANN", The 29th Annual Conference of the Japanese Society for Artificial Intelligence, 29, 1-4, 2015. (In Japanese)

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 327-333 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Improving the Performance of Hadoop Framework Using Optimization Process in the Information Management

Ramachandran Ravi Sowmiyasree, Nachimuthu Maheswari*, Manickam Sivagami

School of Computing Science and Engineering,Vellore Institute of Technology, Chennai,India

ARTICLE INFO

Article history:
Received: 01 June, 2019
Accepted: 17 September, 2019
Online: 22 October, 2019

Keywords:
Bigdata
Hadoop
Performance
Map reduce
Hive
sqoop

ABSTRACT

Hadoop has certain issues that could be taken care to execute the job efficiently. These limitations are due to the locality of the data in the cluster, allocation of the jobs, scheduling of the tasks and resource allocations in Hadoop. Execution in the mapreduce remains a challenge in terms of efficiency. So, an improved Hadoop architecture that takes care of the computation time has been discussed. The improved architecture addresses the communication issues with the task trackers, inefficient clean up task, heartbeat function. Comparing with native Hadoop, the improved Hadoop reduces the total time taken for running the reducer tasks. The performance of the improved system using optimization serves better in the computation time.

1. Introduction

Big data has a huge volume, velocity and variety of information, which produces cost effective, new forms of information processing that helps in decision making, prediction and automation etc. Big data analytics is the process of analyzing large and varied data sets i.e., big data to discover hidden patterns, unknown correlations, trends, user preferences and other useful information that can help organizations to take better decisions in business. Parallel processing in data analytics has appeared as an interdisciplinary research area due to the nature and large size of data [1]. Pattern matching/mining or analysis need huge amount of complex data processing and computing [2].

Significant and reasonable allocation of resources are required to solve complicated problems [3]. It is very difficult to understand and process the data using traditional processing techniques. Big data parallel processing platforms has selected new possibilities to process the structured, semi-structured or unstructured data [4].

1.1 Problem Statement

Data localization and the resource allocation for the tasks are the challenges in the Hadoop. Effective and efficient resource allocation are the challenges in the map reduce framework. In order to implement an architecture with the data gathered from different sources of the systems as in the banking practice and various data processing supports as in the traditional global financial systems, an improvement in the computation time is required.

1.2 Motivation

Financial sectors and the banks are facing severe demands due to the growth of data processing. This happens not only from the improved regulatory requirements and an inconsistency in the data sources. The cost has to be reduced without compromising scalability and flexibility. In this scenario, the financial services industry shows tremendous interest in applying big data technologies to extract the significant value from the large amount of generated data. So, there is a requirement to improve the computation time of the process.

The proposed system addresses the issues in communicating with the task trackers, inefficient cleanup tasks, heartbeat function. Time taken for the reducers to complete the task has been reduced, which helps in better computation time.

The sections of the paper are divided as follows: Section 2 discussed some related works. In section 3, Hadoop map reduce

* Nachimuthu Maheswari, Email: maheswari.n@vit.ac.in

work flow and its performance is discussed. In section 4, the enhanced hadoop system is discussed. In section 5, the implementation phase is discussed. In section 6, the results are evaluated and discussed. In section 7, the conclusion is provided. Finally, in section 8, the future works are discussed.

2. Related Work

The improvement in Hadoop Mapreduce performance has suggested from different concerns or aspects. Many works have suggested in the improvement of the performance of mapreduce jobs and the hadoop, such as scheduling the jobs and computation time improvement. Scheduling the job and the execution time are considered as the important features of Hadoop [5] [6]. Various studies have given the information and have received the results using their ideas [7] [8]. Few people focus on the time of starting and ending process of map reduce job processing [9]. Issues in the system memory could be addressed to improve the performance of the overall system [10]. Apart from Hadoop, few studies utilize a distributed caching method to enhance the performance of Hadoop [11] [12]. Shm Streaming [13] proposes a streaming type of schema to give First In First Out queue as lock less which joins Hadoop and other programs. Hadoop spreads the redundant data into various nodes in various racks. This will be helpful in false tolerant issues. Various studies consider the improvements in data locality development for the betterment of hadoop [14]. Few, discuss on the data types for the betterment in the performance of Hadoop [15].

3. Hadoop Overview

Hadoop is an open source framework for distributed storage and processing. It gives solutions for big data processing and analysis. Hadoop has a file system that stores the data relevant to the applications. The interface is known as the Hadoop Distributed File System (HDFS). Hadoop Distributed File System shares the resources for data analysis. The major parts of Hadoop are map reduce and Hadoop Distributed File System. Other hand, moving the computing towards the data is less costlier than movement of data towards computing [16]. Hadoop allocates file system to store huge amount of data files across the nodes in the clusters.

In HDFS, the Hadoop cluster has two components, which are the masters- Name Node and the slaves- Data Nodes. In Hadoop cluster, name node is responsible for maintaining the file system. It helps in maintaining the data and sending the jobs to the corresponding data nodes responsible for the application data [17]. Job tracker works in the master node. Job Tracker takes care of the data related to the applications in the data nodes with the help of the task tracker for processing. Each task is running in the respective allotted slot in a data node, which has a fixed amount of map/reduce slots.

3.1 Mapreduce

Map reduce processes the job that splits the input data of the job into independent parts and stores the data in HDFS. During map reduce execution phase, multiple map tasks are processed in parallel. After the completion of the map tasks multiple reduce tasks are processed in parallel [18]. With respect to the applications the total number of allotted map tasks be different than that of reduce tasks. Data that are stored in HDFS and processed in map/reduce framework is the form of key and value pair. It has stored and used in the map/reduce tasks to determine the required results at the end of the job. The final results will be displayed in the key and value combination.

3.2 Mapreduce Work flow

HDFS stored the data required for the map reduce job and the data is distributed in to the blocks. Each mapper processed one block in the same time. Inside the mapper phase, the user can give the logic as per theproblem requirements. So that the map tasks runs on all the nodes in the cluster and computes the data parallelly that are stored in the blocks.

The mapper output is stored in the local disk as it is only the intermediate output. If it is written on HDFS, that will create unnecessary multiple copies as the HDFS always replicates the data. The mapper output is sorted/shuffled and submitted to the reducer. The shuffling/sorting has taken care by the internal process. Reducer phase is the next phase of processing and the user can specify his/her own statements. Input to the reducer is given by all the mappers. The final output is produced by the reducer and written on HDFS.

4. Improved Hadoop System

The focus is on transferring the relational data into HDFS using sqoop. The exported data is analysed using Hive by using Hive Query Language [19]. The data is partitioned and is fed into MapReduce for further processing. The MapReduce jobs are optimized due to the changes done in the configuration of Job in progress, Task in progress and Task tracker files of hadoop.

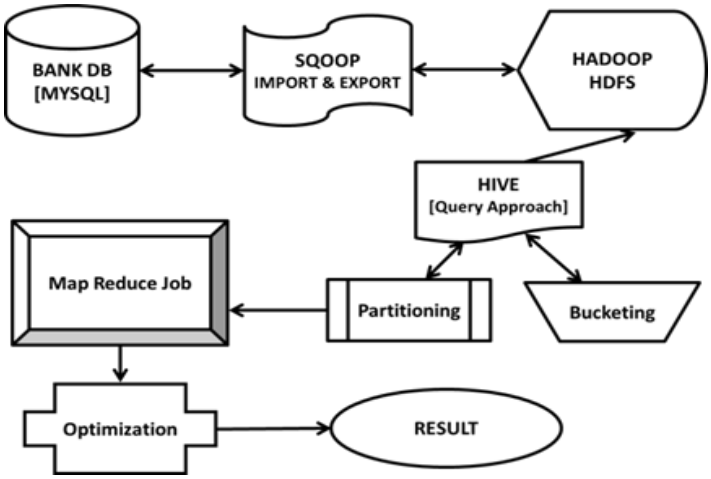


Figure.1 Diagrammatic representation of the proposed system

The bank MySql Database contains the required datasets. The data is exported to HDFS using Sqoop tool. Then Hive does the partition of the exported data, for which the results gets stored in HDFS. Then the Hive partitioned data is analysed and fetched from HDFS for MapReduce processing. The next step is Optimisation, which is like the heart of the work. Once the optimisation is over, the results are compared with that of the native hadoop to prove the working correctness of the work as

shown in Figure 1. The flow of the proposed system is explained below:

System Work flow

Step 1: Client requests for data sets. Bank data set is transferred from Mysql to HDFS using Sqoop.

Step 2: Hive extracts the data and partition the data set for map reduce programming.

Step 3: Changes performed in the Jobinprogress, Taskin progress, task tracker files.

Step 4: Name node initializes the job. Data fetched from the data node and performs the job.

Step 5: Performance of the results with respect to the computation time is compared with the native Hadoop.

4.1 Data Collection and Preprocessing

The Initial phase of the work is data collection. The article has a bank dataset that contains the account details, customer details and transaction details in MySql database.The transfer of the dataset into hadoop (HDFS), i.e, data migration is done using Sqoop tool. Sqoop supports for transferring data from relational databases to Hadoop.

Data from the three datasets (account,customer and transaction) are combined with its attributed and produced 21,215 records. The final combined dataset has the following attributes: account number, transaction type, amount, month and year.

In this phase, the dataset is fetched into hadoop (HDFS) using Sqoop Tool as in Figure 2. Using Sqoop, customized functionalities can be performed like fetching the particular column or fetching the dataset with specific condition and data can be stored in hadoop (HDFS).

Figure.2 Data preprocessing using sqoop

4.2 Data Analysis

In this phase, the dataset is analysed using HIVE tool which will be stored in hadoop (HDFS). For analyzing the dataset, HIVE uses HQL Language. Hive is chosen because it is a data warehouse by itself. Hive has the ability to work on top of an existing Hadoop cluster and provides SQL-like interface. The user can map the existing Sqoop tables to Hive and operate on them as shown in Figure3.

4.2.1 Partitioning

Hive arranges the tables into partitions. The records are partitioned based on the transaction type(with draw, transfer, deposit). Partitioning helps in dividing a table into related parts based on the values of partitioned columns as shown in Figure 4. Partitioning is useful in querying a portion of the data.

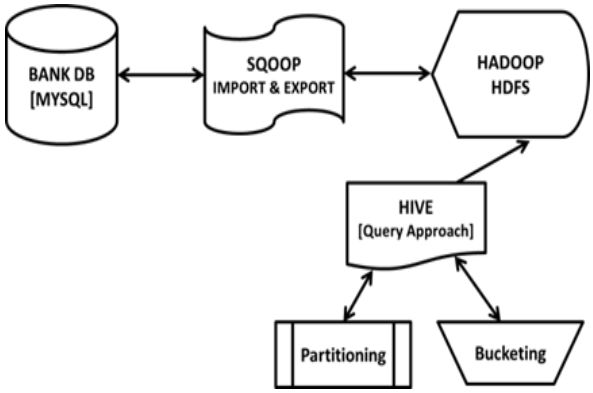


Figure.3 Data analysis using Hive

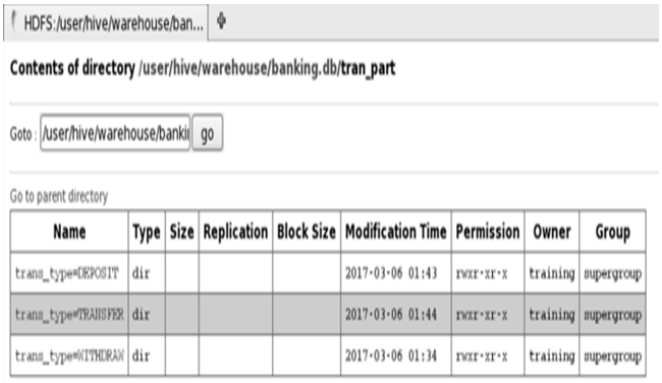


Figure.4 Partitioning the transaction data

The number of rows loaded and the time taken to load the partitioned data into HDFS is shown in Figure 5.

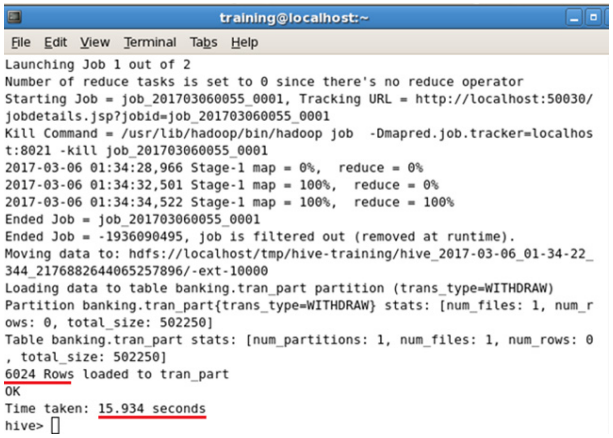


Figure.5 Loading partitioned data into HDFS

4.2.2 Bucketing

Bucketing decomposes data into manageable or equal parts as shown in Figure 6. The user can restrict the number of buckets to store the data. It provides faster query response. The number of required buckets can be given during the table creation. Loading of equal volume of data has to be done manually by programmers is a significant issue.

Choosing partitioned data over bucketed data

The size of the loaded partitioned data in the three partitions is less when compared to the size of the loaded bucketed data.

Bucketing works well when the field has high cardinality (number of possible values a field can have) and data is evenly distributed among buckets. Partitioning works best when the cardinality of the partitioning field is low.

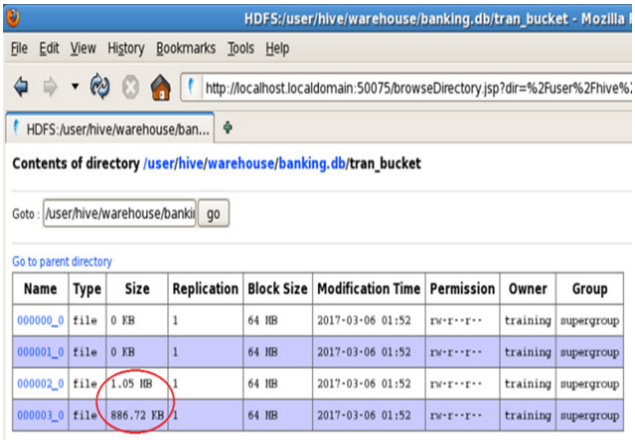


Figure.6 Bucketed data in HDFS

4.3 Data Processing

Data Processing is done using Hadoop Madpreduce. The partitioned data is fed into MapReduce for the processing. The main factor considered is the SLOTS_MILLIS_REDUCES[20]. It represents the total time taken for running the reduce tasks in milliseconds in the occupied slots. It also includes the tasks that where started speculatively as shown in Figure 7.

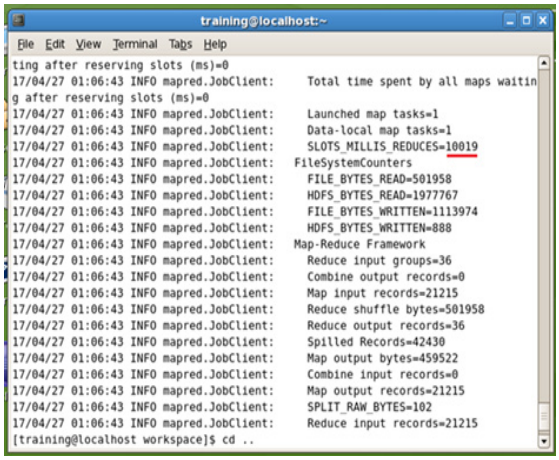


Figure.7 Running mapreduce job

4.4 Optimisation

Optimisation can be done to the native hadoop by changing the configuration factors of the system. There are three main files in hadoop[21], to which the changes are done to obtain an improved and an optimised version of hadoop. The files in hadoop to which the changes are made in order to obtain an improved version are

JobInProgress.java

JobInProgress maintains all the information in order to keep the job on the straight. It maintains its JobProfile and its updated JobStatus, along with a set of tables for doing bookkeeping of its tasks.

TaskInProgress.java

TaskInProgress (TIP) maintains all the information required for a task in the lifetime of its owning Job. A given task might be executed or re-executed, so level of indirection is needed above the running-id itself. A given TaskInProgress contains multiple taskids, zero or more of which might be executing at any one time, allowing speculative execution. A TIP allocates enough taskids to account for all the speculation and failures it will ever have to handle. TIP is dead when they are up.

TaskTracker.java

Task tracker is a process that initiates and monitors the map reduce tasks in the environment. It communicates the Job tracker for assigning the tasks and update the status.

4.4.1 JobInProgress

The changes made to the JobInProgress.java are as follows,

Contacting the tasktracker at once

In native hadoop, the task can contact the TT only after the whole task gets completed. So change is made so that, once the runJobSetupTask() works, the task can contact the TT then and there. It is beneficial as the TT is well informed of the on-going tasks and can also help to avoid speculations.

Launching Synchronised Cleanup

In native hadoop, Cleanup task is called once at the end of the task. The change made here is the introduction of Synchronised Cleanup task, which is launched before the cleanup task. It updates the last known cluster size then and there. It is highly beneficial in case of multiple tasks.

Later check of Cleanup

It checks whether the cleanup task is launched already or if the setup is not launched already. The later check is beneficial when the number of maps is zero.

4.4.2 TaskInProgress

The changes made to the TaskInProgress.java are as follows,

Skipping feature

Since completed reduces (for which the outputs go to hdfs) are not failed, this failure is noted only for completed maps, only if the particular tasked completes the particular map. However if the job is done, there is no need to manipulate completed maps. Reset the successful TaskId since no successful tasks can be concluded while running. There can be failures of tasks that are hosted on a machine that has not yet registered with restarted jobtracker. Therefore, the skipping feature recalculates the counts only if it is a genuine failure.

SetComplete function

When the TIP is complete, the other speculative subtasks will be closed when the owning tasktracker reports in and calls should close on the required object.

4.4.3 TaskTracker

The changes made to the TaskTracker.java are as follows,

Resetting heartbeat interval from the response

The name node and the data node communicate using Heartbeat messages. After a certain amount of time, if the Name Node does not receive any response from Data Node, then that particular Data Node is declared as dead. The normal heartbeat period is three seconds. If the name node is not receiving the heartbeat from a data node in ten minutes, then the name node decides the data node to be not in service and the data node's replicas are considered to be unavailable. The name node allocates the new replicas for the unavailable blocks in the other data nodes. The heartbeat message from the data node carries the information regarding the storage, amount of usage and transfer rate of data in the current state. Such details are helpful for the name nodes allocation of blocks and the decision making in load balancing.

By resetting heartbeat interval from the response, the TT knows the completed tasks. Once cleanup of the completed task is over, normal operation is resumed.

5. Implementation

5.1 Mapper Class

The map task is to process the input data. The input data is stored in the Hadoop Distributed File System (HDFS). The input file is passed to the map function and read line by line. The mapper processes the data and produces the intermediate data. The amount, transaction type, year and month are initialised according to the positions in the input data. The mapper class writes the year, month, transaction type and amount along with the sum into HDFS.

5.2 Reducer Class

This stage is the combination of the shuffle and the reduce stage. The Reducers job is to process the data that comes from the mapper. After processing, it produces a new collection of output, which will be stored in the HDFS.

5.3 Driver Class

The Driver class verifies the arguments from the command line. The input/output file details, job details are available in the command line. It assigns values for the job. The driver starts the execution from the main () method. In this method, a new configuration object and the Job are instantiated.

The job.waitForCompletion() helps to unveils the job perfectly. This driver code helps to wait for the job completion. The job status will be updated after the job completion. The true argument informs the framework to write verbose output to the controlling terminal of the job.

6. Results and Discussion

The work has done using bank data, which is analysed to identify the future trends in the sector. The improvement in the Hadoop performance is ensured by the factors such as computation time, time taken for running reduce tasks, contacting the task trackers at once, synchronised clean up, skipping feature, resetting the heartbeat interval.

6.1 Health of the file in HDFS after optimization

The fsck command in hadoop runs a HDFS file system checking utility. It is designed for reporting problems with various files. It checks health of hadoop file system. It produces a report that gives the complete health of the file system. HDFS is considered healthy if all the files have a minimum number of replicas as shown in Figure.8.

Table 1. Performance Analysis

Factors	Before Optimization(sec)	After Optimization(sec)
Job Duration	82	23
Map Time	60	3
Reduce Time	22	20

```
[training@localhost workspace]$ hadoop fsck /user/training/after_optimisation -f
iles -blocks -locations
FSCK started by training from /127.0.0.1 for path /user/training/after_optimisat
ion at Wed May 03 02:16:39 PDT 2017
/user/training/after_optimisation <dir>
/user/training/after_optimisation/_SUCCESS 0 bytes, 0 block(s):  OK

/user/training/after_optimisation/_logs <dir>
/user/training/after_optimisation/_logs/history <dir>
/user/training/after_optimisation/_logs/history/localhost_1490078156466_job_2017
03202335_0037_conf.xml 47417 bytes, 1 block(s):  OK
0. blk_-5486851598621712856_2862 len=47417 repl=1 [127.0.0.1:50010]

/user/training/after_optimisation/_logs/history/localhost_1490078156466_job_2017
03202335_0037_training_bank.jar 9391 bytes, 1 block(s):  OK
0. blk_6501123563346896786_2865 len=9391 repl=1 [127.0.0.1:50010]

/user/training/after_optimisation/part-r-00000 888 bytes, 1 block(s):  OK
0. blk_3551654415947294813_2864 len=888 repl=1 [127.0.0.1:50010]

Status: HEALTHY
 Total size:    57696 B
 Total dirs:    3
 Total files:   4
 Total blocks (validated):      3 (avg. block size 19232 B)
 Minimally replicated blocks:   3 (100.0 %)
 Over-replicated blocks:        0 (0.0 %)
 Under-replicated blocks:       0 (0.0 %)
 Mis-replicated blocks:         0 (0.0 %)
 Default replication factor:    1
 Average block replication:     1.0
 Corrupt blocks:                0
 Missing replicas:              0 (0.0 %)
 Number of data-nodes:          1
 Number of racks:               1
FSCK ended at Wed May 03 02:16:39 PDT 2017 in 2 milliseconds


The filesystem under path '/user/training/after_optimisation' is HEALTHY
```

Figure.8 Health status of the result

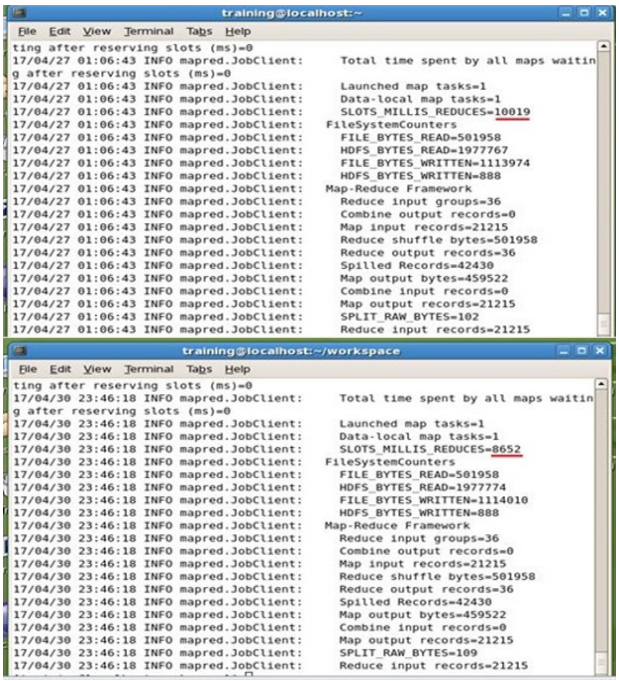


Figure.9 Performance - before and after optimization

SLOTS MILLIS REDUCES is the total time spent by all reduces in the occupied slots. Difference in the time taken for running reduces tasks before and after optimisation is noted to be 1367ms as shown in Figure 9. After optimization it takes less time.

6.2 Performance Analysis

The performance of the job for the same input is analysed before and after the optimisation is show in Table 1. Job duration, Map task time and Reduce task time are taken as the factors for the performance analysis. The time taken to complete the job is less after optimization is shown in Figure 10. The performance has evidently improved after the optimisation.

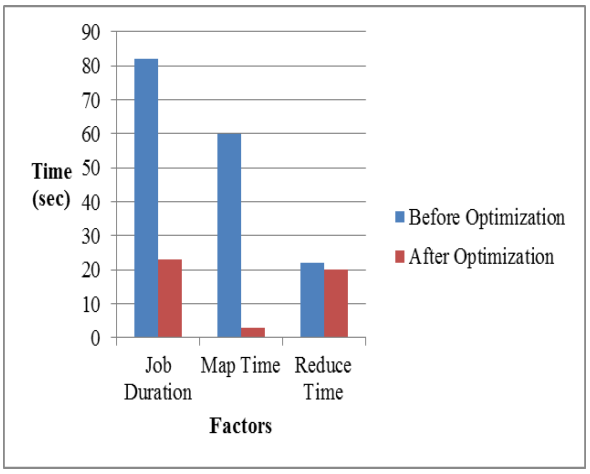


Figure.10 Performance Evaluation

7. Conclusion

In this article, an improved hadoop framework is presented. It includes analysing the internal working of the Hadoop MapReduce job, so that changes are done to the job in progress, task in progress and task tracker files. In H2Hadoop [19] the work was concentrated on the DNA dataset only and in the improved hadoop [22] they have considered the text data and the performance time was not reasonably improved. The changes performed to optimise the system in this article are in contact with the job tracker at once, launching synchronised cleanup function, cleanup later checkup, skipping feature, resetting the heartbeat interval from the response, avoiding speculative tasks resulted in an improved hadoop framework. The performance of the improved system using optimization is better in the computation time.

8. Future Work

Performance factors

The other performance factors of File System Counters, MapReduce Framework can be modified to obtain a more optimised hadoop system.

Tool for better performance

The result can be analysed using Spark to obtain a faster performance than Hadoop, as spark runs in-memory on the cluster, and it is not tied to Hadoop MapReduce paradigm. This makes repeated access to the same data much faster. Spark can run as a standalone or on top of Hadoop YARN, where it can read data directly from HDFS.

Security issues in Hadoop can be addressed at various levels, including but not limited to file system, networks, scheduling, load balancing, concurrency control, and databases.

References

[1] MMing, M., G. Jing, and C. Jun-jie, "Blast-Parallel: The parallelizing implementation of sequence alignment algorithms based on Hadoop platform", *In: Proc. of International Conf. on Biomedical Engineering and Informatics (BMEI)*, IEEE, pp. 465-470, 2013. doi**:** 10.1109/BMEI.2013.6746988

[2] cC. Schatz, B. Langmead and S. L. Salzberg, "Cloud computing and the DNA data race", *Journal of Nature biotechnology*, Vol.*28, No.*7, p.691, 2010. doi:10.1038/nbt0710-691.

[3] E. E. Schadt, et al., "Computational solutions to large-scale data management and analysis", *Journal of Nature Reviews Genetics*, Vol.11, No.9, pp. 647-657, 2010. https://doi.org/10.1038/nrg2857.

[4] Marx, *Biology: The big challenges of big data*. Nature Publishing, 498(7453): pp. 255-260, 2013. https://doi.org/10.1038/498255a

[5] N. Tiwari, S. Sarkar, U. Bellur and M. Indrawan, "Classification framework of MapReduce scheduling algorithms", *International Journal of ACM Computing Surveys (CSUR)*, Vol.47, No.3, p.49, 2015.doi:10.1145/2693315

[6] M. Zaharia, et al., *Job scheduling for multi- user mapreduce clusters*, Tech. Rep. UCB/EECS-2009-55, EECS Department, University of California, Berkeley, 2009. doi =10.1.1.649.3097

[7] R. Gu, et al., "SHadoop: Improving MapReduce performance by optimizing job execution mechanism in Hadoop clusters", *Journal of Parallel and Distributed Computing*, Vol.74, No.3, pp. 2166-2179, 2014. https://doi.org/10/1016/j.jpdc.2013.10.003

[8] Q. Chen, C. Liu, and Z. Xiao, "Improving MapReduce performance using smart speculative execution strategy", *Journal of IEEE Transactions on Computers*, Vol.*63, No.*4, pp.954-967, 2014. doi:10.1109/TC.2013.15

[9] Y. Jinshuang, et al., "Performance Optimization for Short MapReduce Job Execution in Hadoop", In: *Proc. of International Conf. on Cloud and Green Computing (CGC)*, IEEE, pp. 688-694, 2012. doi:10.1109/CGC.2012.40

[10] Apache, Centralized Cache Management in HDFS. Update date 2014.https://hadoop.apache.org

[11] S. Zhang, et al., "Accelerating MapReduce with distributed memory cache", In: *Proc. of International Conf. On Parallel and Distributed Systems (ICPADS),* IEEE, pp.472-478, 2009. doi:10.1109/ICPADS.2009.88

[12] J. Zhang, G. Wu, X. Hu and X. Wu, "A Distributed Cache for Hadoop Distributed File System in Real-Time Cloud Services", *In: Proc. of International Conf. On Grid Computing,* IEEE Computer Society, pp.12-21, 2012. doi:10.1109/Grid.2012.17

[13] L. Longbin, et al., "ShmStreaming: A Shared Memory Approach for Improving Hadoop Streaming Performance", In: *Proc. of International Conf. On* Advanced Information Networking and Applications (AINA), IEEE, pp.137-144, 2013. doi:10.1109/AINA.2013.90

[14] B. Palanisamy, et al., "Purlieus: locality-aware resource allocation for MapReduce in a cloud.", *In: Proc. of International Conf. On High Performance Computing, Networking, Storage and Analysis*, ACM, pp.1-11, 2011. doi:10.1145/2063384.2063462

[15] H. Alshammari, J. Lee and H. Bajwa, "H2Hadoop: Improving Hadoop Performance Using the Metadata of Related Jobs," in *IEEE Transactions on Cloud Computing*, vol. 6, no. 4, pp. 1031-1040, 2016.doi:10.1109/TCC.2016.2535261

[16] Hammoud, and M. F. Sakr, "Locality-aware reduce task scheduling for MapReduce", *In: International Conf. on Cloud Computing Technology and Science (CloudCom),* IEEE, pp. 570-576, 2011. doi:10.1109/CloudCom.2011.87

[17] J. B. Buck, et al., "SciHadoop: Array-based query processing in Hadoop", *In: Proc. of International Conf. on High Performance Computing, Networking, Storage and Analysis (SC)*, 2011. doi:10.1145/2063384.2063473

[18] Holmes, *Hadoop in Practice*, Manning Publications Co., 2012.

[19] Edward, W. Dean and R. Jason, *Programming hive*, O'Reilly Publications, 2012.

[20] Y. Xiao and B. Hong, "Bi-Hadoop: Extending Hadoop to Improve Support for Binary-Input Applications", In: *Proc. of International Symposium On* Cluster, Cloud and Grid Computing (CCGrid), IEEE/ACM, pp.245-252, 2013. doi:10.1109/CCGrid.2013.56

[21] https://hadoop.apache.org/docs/r2.7.1/api/org/apache/hadoop/mapreduce/JobCounter.html

[22] Thanekar, Sachin & Bagwan, A. & Subrahmanyam, K..,Improving Hadoop performance by enhancing name node capabilities.Fronteiras. 6. 1-18.,2016. doi:10.21664/2238-8869.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 334-340 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

A Modular Multipurpose, Parameter Centered Electronic Health Record Architecture

Norbert Maggi[1,2], Luca Douglas Magnoni[2], Carmelina Ruggiero[1, 2], Roberta Gazzarata[2], Mauro Giacomini[1, 2,*]

[1]*Department of Informatics, Bioengineering, Robotics and Systems Engineering (DIBRIS), University of Genova, Genova, Italy*

[2]*Healthropy S.r.l., Savona, Italy*

ARTICLE INFO

Article history:
Received: 15 June, 2019
Accepted: 06 September, 2019
Online: 22 October, 2019

Keywords:
EHR
HL7
Interoperability

ABSTRACT

Health Information Technology is playing a key role in healthcare. Specifically, the use of electronic health records has been found to bring about most significant improvements in healthcare quality, mainly as relates to patient management, healthcare delivery and research support. Health record systems adoption has been promoted in many countries to support efficient, high quality integrated healthcare. The objective of this work is the implementation of an Electronic Health Record system based on a relational database. The system architecture is modular and based on the concentration of specific pathology related parameters in one module, therefore the system can be easily applied to different pathologies. Several examples of its application are described. It is intended to extend the system integrating genomic data.

1. Introduction

This paper is an extension of work originally presented in IEEE 4th International Forum on Research and Technology for Society and Industry [1]

The collection of health information for patient data management including patient characteristics, treatments and outcomes has been over the years a most important aspect of health care [2, 3]. EHRs, which replace paper medical records to collect people's health information, integrate different and heterogeneous data and are one of the most widely used Health Information Technology (HIT) tools. HIT contributes to the improvement of quality of care, to patient's safety and to the reduction of health care costs by making available accurate documentation and rapid information retrieval and management. The adoption of EHRs has been promoted in many countries, both at government level an in the private sector, even though their adoption is still rather limited, mostly because of standardization, certification, security and privacy aspects and concerns. The adoption of standard based EHR systems has been found to be able to reduce spending for healthcare to a great extent [4]. EHRs can improve patient care and safe practice, however potential risks related to medical error, systems failure, and security and legal responsibility aspects should be considered, identifying barriers for EHRs adoption and ways to address them [5-9]. Moreover, organizational and human factors can promote or hamper the adoption of a EHR [10]. Further relevant aspects of EHRs are interfaces with other information systems, that should be based on standard formats, such as Health Level 7 Clinical Document Architecture (HL7 CDA), HL7 Version 3, openEHR and on controlled vocabularies such as Logical Observation Identifiers Names and Codes (LOINC), "International Classification of Diseases (ICD), Systematized Nomenclature of Medicine (SNOMED) and others [11]. The need for generic and interoperable EHRs has been recognized, specially as relates to interoperability, for the preservation of clinical information across heterogeneous systems [12]. Standards are needed to implement interoperability between systems.

An EHR system contains quantitative data (clinical data), qualitative data (text documents), and medication records. The related amount of data is rapidly expanding, and big data technology is going to play a key role in this respect, specially as relates to natural language processing, clinical decision support, integration of genomics and system biology with EHR data, knowledge dissemination and interaction with patients [13].

The set of requirements that must be met by EHR systems is defined by ISO 18308:2011. According to this standard the scope of the EHR is regarded as broader than documentation, prevention and treatment of illnesses. ISO 18308 relates to shared EHR information and to the EHR system aspects that may be used for the support and coordination of patient-centred continuity of care. An EHR system includes data repositories, directory services,

*Corresponding Author: Mauro Giacomini. Via all'Opera Pia, 13 – 16145 Genoa (Italy) email: mauro.giacomini@dibris.unige.it

knowledge services containing terminological system, care pathways and workflows, user applications and other modules and services. An EHR architecture can be classified according to a data approach, a concepts approach and a process/service approach. These approaches are present in existing systems and in specific domain languages and modelling languages [14]. A medical record system has been developed, adopting the three approaches above. The medical record system that has been developed is based on Health Level 7 (HL7) standard and can be easily tailored to the needs of different clinical wards. The use of this standard allows the system to support interoperability between the main components of the Hospital Information System (HIS), such as the Laboratory Information System (LIS) and the Pharmaceutical Information System (PIS), and can be easily extended for example to the radiology information system (RIS) and to the picture archiving and communication system (PACS). A Service Oriented Architecture (SOA) has been adopted, that can be defined as an open extensible, composable architecture based on autonomous, interoperable and reusable services, implemented as a Web services [15]. SOA has been employed in healthcare after adoption in other sectors, such as banking, transport and industrial automation [16]. SOA is widely used for real time implementation in large distributed systems, mostly because it supports the easy integration of new software within the existing one, reducing the impact on service users and cost [17-20]. Moreover, the SOA paradigm has been successfully implemented in the design of several distributed healthcare systems [21-23].

2. Materials and Methods

The system has been implemented taking into account the specific needs of different departments of the San Martino hospital in Genoa (Italy).

The workflow relating to a visit includes the identification of the patient through the information already present in the HIS obtained by a pseudo-anonym code according to the specifications of UE General Data Protection Regulation, (GDPR) [24]. The information on laboratory analyses is obtained by query of the hospital LIS and the information on therapies is obtained by query of the PIS of the hospital. The data are stored on different platforms which do not interact with each other, therefore its direct use is cumbersome and with possible misunderstandings. In order to address this problem, a system based on a relational database developed with Microsoft SQL Server 2017 has been set up, which has made it possible to create a consistent and logical representation of the information. It is worth noting that, in order to achieve GDPR compliance, the data in the database are stored by pseudo-anonymization techniques. A pseudo-anonymous code is used to query the database, in a secure fashion. The information relating to the identity of the patient is present only at the physician interface level, after authentication [23].

The database has been designed developing a conceptual model of the data, abstracting relevant aspects of the activities and processes relating to patient management. A conceptual model is a set of rules and conventions that allows to move from a reality of interest to a conceptual scheme of the data. The model has been divided into three levels that represent respectively the protagonists of the project (level 0), the clinical event (level 1), and the format of the parameters and results (level 2).

In level 0 the main actors and features of the project have been considered, defining the fundamental entities and the relationships among them. The Entity-Relation (E-R) diagram shows the general logistical aspects of the platform (Figure 1). Patients are followed up and subsequently entered into the database by the operators. Both patients and operators are linked to the centres involved in each clinical study. Each operator may be authorised to access more than one clinical study, and each patient can be involved in more than one study. Moreover, the clinical events relating to each patient may take place in different centres (the latter aspect being of fundamental importance, since different centres may use different units of measurement and ranges of normality). Each study may consider more than one clinical event, some of which may be common to those of other studies, while others may be specific to certain studies only.

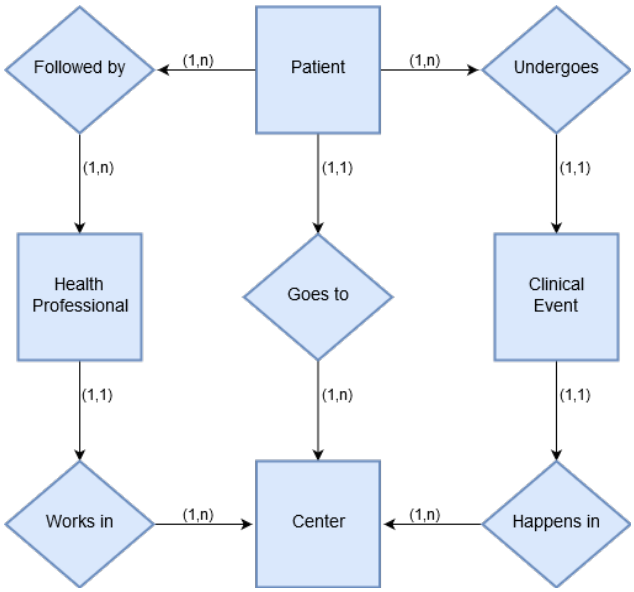


Figure 1. Entity relationship diagram for level 0

Level 1 of the E-R diagram (Figure 2) shows the meta-descriptive approach that has been adopted to manage clinical events and make the database structure as simple and general as possible. The main element is the definition of different Event Types. Within the conceptual definition of systems implemented with E-R diagrams [25], attributes are the elements that describe in detail the characteristics of entities and relationships. By transforming E-R diagrams with relational logic schemes, these attributes become columns of tables in databases. This approach, if applied to our case, binds the set of elements considered in a medical record (which happens in many third-party applications), blocking it in the data base (DB) construction.

When developing the system, it has been chosen to store the names of the elements that are regarded as rows of the table (called "parameters"). Another table (called "categories") has been categorized, and related with the clinical reality through the links in the table that derive from the relation shown in in figure 2 (diamond "considered aspect divided by"). This choice allows to use the same DB and the same software for several tables and to easily characterize the individual parameters in a strong semantic way (using international codes such as - LOINC - ICD - ...). This greatly helps interoperability.

The most important aspects have been divided into different "Categories". These are "containers" of parameters, homogeneous

from the conceptual point of view, but not necessarily homogeneous from a data format point of view (for example, a clinical event of the "Therapy" type has a "Therapy information" category, including parameters such as "Adherence" and "Therapy start date"). Each clinical event has Results, which are the values of the parameters of that particular event.

Events are grouped into types. Parameters are also grouped into types, which are translated and mapped in LOINC using standardized dictionaries, in order to support data reuse.

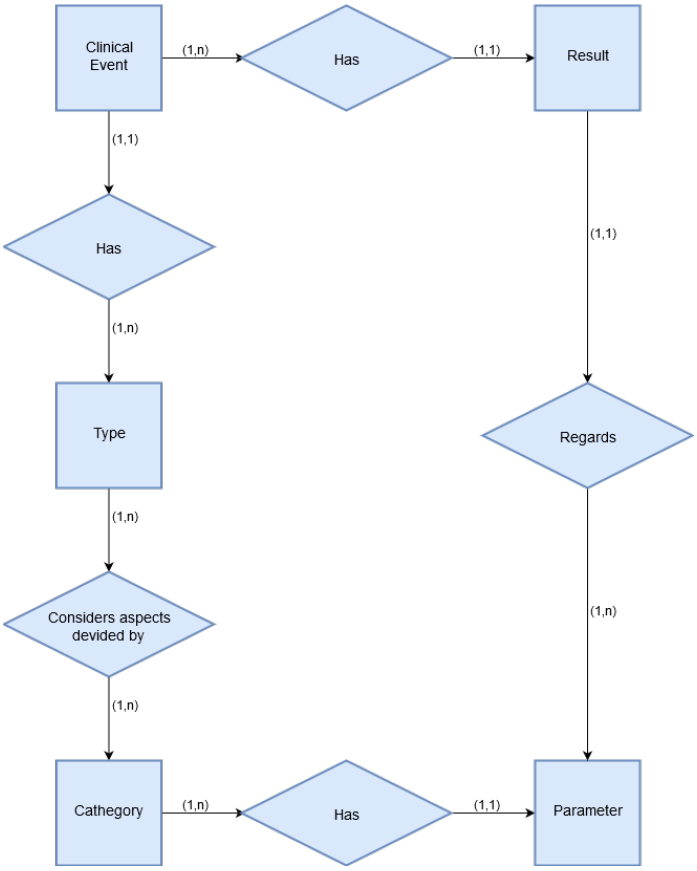


Figure 2. Entity relationship diagram for level 1

Level 2 shows the entities and relationships which are taken into account to manage the Parameters and Results of Clinical Events. Each Parameter has a specific format (typically real integer, string or Boolean), so the results (i.e. the records corresponding to the values of the parameters of an event) can have different formats. The E-R diagram (Figure 3) takes into account the fact that all parameters stored in the database have their own units of measurement and range of normality, but these units and ranges are different for each centre that may be involved in the network. This is of fundamental importance for the work carried out by the physician. In the transition from the conceptual scheme to the logical scheme, great attention has been paid to this aspect, in order to allow each user to work with the tools that are most familiar to him/her. Moreover, the differences between the units of measurement have been taken into account in the data extraction algorithm. All records relating to a single parameter are converted into a standard unit of measurement, defined by the users as the ones which are most widely used.

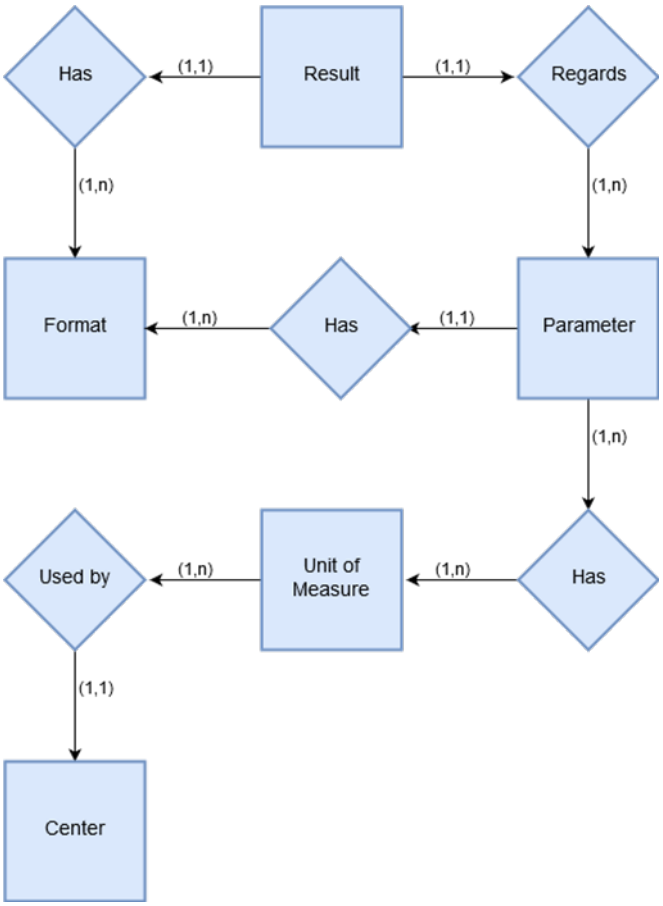


Figure 3. Entity relationship diagram for level 2

The implementation, including the creation of a web client and a web service, has been carried out in Visual Basic .NET using Microsoft Visual Studio 2015 software.

The system architecture has been designed with special reference to interoperability, according to HL7 in the HSSP (Healthcare Services Specification Project). The HSSP, started in 2005 by HL7, focuses on standard software services for healthcare, enabling the use of SOA, which has been recognised as extremely useful for interoperability, code reuse and new needs response [26]. Standards service interfaces for infrastructure reuse capability and specific health facility were considered providing a methodology for the implementation of relevant services [27].

Specifically, the HL7 Clinical Document Architecture (CDA) has been used for the encapsulation of clinical data that can be easily exchanged between applications, retaining their semantic significance [28].

Clinical parameters are associated differently with specific names and internal codes in different hospital and/or health centres. In order to allow the communication among different centres, the codes used by different hospitals/centres have been translated into LOINC language with regard to clinical trials [29], international Anatomical Therapeutic Chemical Classification System (ATC), national Marketing Authorisation (AIC) for drugs, International Classification of Diseases 9th version (ICD9) for diagnosis and ISO Country 3166-1 for the classification of the country of origin of patients.

Table 1. Clinical events for the Maculopathies, HIV, and Tuberculosis

Maculopathies	**HIV**	**Tuberculosis**
Personal data	Blood sample taken	First visit
Anamnestic	Admission history and physical note	Visit
Calendar	Medical history unknown	Microbiological examination
Eye examinations	Therapy	Laboratory examination
Maculopathies	Discharge summarization note	Resistance development
Laser Injection Therapies	Specialist visit	Instrumental examination
Primary Diagnosis	Submission questionnaire	Therapeutic treatment
Secondary Diagnosis	Non-Haart Therapy	Alcohol test
Clinical Folder	Adverse event	
Two Eyes	Qualitative therapy	
Pharmacovigilance Card	Repeated medical history	
Ozurdex Monitoring Card	Vaccination	
Visual Functioning Questionnaire	Microbiological culture	
Retina Plus - Personal data	Failure	
Retina Plus - Access	Instrumental examination	
Retina Plus - Questionnaire	Visit Control	
Telescreening	Pregnancy	
Laboratory Exams		
Plaquenil Protocol		
Certificates		
Allergies		

The “San Martino” hospital in Genoa is equipped with a network infrastructure made of two Local Area Networks (LANs) that are not connected to each other (the hospital LAN and the university LAN). Specifically, the parts of the HIS are placed in the hospital LAN, which is not accessible from external networks for security reasons, because of the hospital firewall. The access to data and to the required information is granted by a virtual private network (VPN), capable to securely connect the two LANs, allowing access for clients belonging to the university network.

The original data are sent through the VPN and, by tunnelling protocols and advanced encryption, are secured by a protective barrier before they are shipped over a potentially dangerous infrastructure such as the Internet. Tunnelling is a multi-protocol (IPsec) data encapsulation process within another IP packet that is shipped over the network. In order to transmit the data to authorized end recipients only, the data are packaged twice by a specific authentication system in which access levels depend on user role.

The strong modularity of the system allows to use the general structure described above for different diseases. The system has so far been used for the implementation of the EHR for the Ligurian HIV network [30-32], for ophthalmology [33, 34] and for infectious diseases [35]. For each of these cases, different clinical events, parameter categories and parameters have been implemented in order to provide a tool which is very suited to the specific needs of physicians and caregivers. Table 1 and table 2 show, respectively, the clinical events and the categories of parameters for the three scenarios.

3. Results

The system that has being developed is hosted in Genoa university network. The user interface of the system is divided into several parts, each concerning a specific section of the medical visit. The EHR system is managed through a web interface that, for security reason, can only be accessed within the Genoese hospital and university LANs. The system interaction with the physician or caregiver takes place as follows. When the user logs in he/she is redirected to the page relating to the specific patient selection and the examination starts. After obtaining access permission, the caregiver can either manually enter the data of a new patient or see the data for a patient who is already present in the data base.

For a new entry, the related information is retrieved from the hospital databases and, after translation into LOINC, is stored in the EHR system database. The clinical parameters are retrieved from the LIS and the information on the laboratory tests is stored by an array of CDA which is sent to the web service which stores it in the system database.

A section of the clinical diary allows physicians to inspect the exams that are available for each patient. The physician can see all specific visits, their dates and parameters. The information for each clinical event is retrieved and updated by algorithms that taking into account the ethical committee permissions.

The creation of the web pages for user interface is also guided by the structure of the DB, i.e. a page is formed for each event, within which all parameters related to that event are presented. The type of value for each parameter (integer, real, list or free text, date) automatically inserts the selection of the most suitable input tool for that type (textbox, dropdown list, calendar, ...). The measurement units for the numerical parameters and the normal ranges also (possibly linked to the patient's personal data, age, sex, height, weight) are also automatically inserted.

Table 2. Categories for Maculopathies, HIV, and Tuberculosis

Maculopathies	**HIV**	**Tuberculosis**
Personal data	Phenotype and viremia	Symptomatology
Anamnestic	Carbohydrate metabolism	Inflammation indices
Calendar	Cardiovascular profile	Diagnosis of tubercular infection
Visuals and symptoms	Liver functionalities	Sputum microbiological examination
Front and rear segment	Renal function	Microbiological examination of bronchoaspirate
Laser Injection Therapies	Oxidative stress	Microbiological examination of urine
Primary diagnosis	NK phenotype	Microbiological examination of faeces
Secondary diagnosis	Anamnestic	Microbiological examination of other material
Topical therapies in progress	Non-Haart Therapies	Resistance
Expert	Inflammation parameters	Instrumental examinations
Folder	Leukocyte formula hemochrome	Localization
Exclusion criteria	Monocyte populations	Inpatient
Prohibited treatments	Mocyte activation markers	Comorbidity
Interruption of treatment	Inflammation mediators	Treatment
Pharmacovigilance Card	Lymphocyte Proliferation	Side effects
Ozurdex Monitoring Card	Therapy information	First-line oral agents
Visual Functioning Questionnaire	Drugs therapy	B: Injectable agents II line
Appendix of Optional Additional Questions	Information about hospitalization	A: Fluoroquinolones
Retina Plus Personal data	Diagnosis of hospitalization	C: Other second line agents
Retina Plus Access	Diagnostic and therapeutic procedures	D1: Deputy agents
Retina Plus Questionnaire	Discharge information	Laboratory examination
Telescreening	Comorbidity	CBC
Grape harvesting	Discharge diagnosis	D2: Deputy agents
Laboratory Exams	Non-Haart Drugs	D3: Deputy agents
Instrumental examinations	Objective examination	
Microbiology tests	Chemistry	
Plaquenil start therapy	Specialist visit information	
Plaquenil	HIV information	
Blood pressure	HBV information	
Certified	HCV information	
Allergies	Hepatitis	
	T lymphocyte activation markers	
	Side effects therapy	
	Submission questionnaire	
	Non-Haart Drugs	
	Information on drugs therapy	
	Details of side effects therapy	
	T reg	
	Concomitant therapy	
	Adverse event therapy	
	HCV adverse event	
	Qualitative HCV therapy	
	Qualitative HIV Therapy	
	Vaccination	
	Microbiology	
	Non-hepatitis serologies	
	Instrumental examination information	
	Control visit information	
	Resistance	
	Pregnancy	
	Barthel ADL (Activities of Daily Living)	
	Barthel MOB	
	IADL	

Table 3. Use of EHRs for pathologies

	Initial year	Patients	Parameters	Events	Values	Sessions
HIV	2013	4540	802	200424	6481673	9033
Maculopathies	2014	5554	801	52335	955753	2166
Tuberculosis	2016	324	125	6460	103691	2324

4. Discussion and Conclusions

An EHR system has been set up in order to meet the needs of medical and clinical personnel. The main aspects that have been considered relate to the management of visits and to the setting up of a clinical diary according to the standards currently in use.

The objectives of the study have been met as shown in table 3, in which the system's application to maculopathies, HIV, and tuberculosis are described. Specifically, for the three pathologies the table shows the initial year, the numbers of patients which have been recorded, the numbers of related parameters, the number of events and the number of values that have been stored. The last column, named "session", shows the number of days in which an operator queries an EHR.

The system aims to simplify physician's, nurses' and caregiver's work for administration and management of treatment and follow-up. Moreover, it aims to support multicentric tracking for patients who are treated in several hospital facilities and the conduction of clinical trials. The system was set up focusing on parameters and parameter types, grouping them in one specific module. This facilitates the use of the system for different pathologies, because the pathology related data are in one module per pathology. The use of one single section for the management of different pathologies allows for easy use by different departments of one hospital and by different hospitals. This design choice also facilitates data reuse.

The main problems and limitation that have been encountered relate to data availability, because of data anonymity requirements for security and privacy. In the first place, this problem has been circumvented requiring the hospital to make the data pseudo-anonym, while keeping the link between the pseudo anonym code and the patient identity separate from the data and managed by the hospital system only, on physically separated platform. Moreover, the data which are copied onto the EHR system platform are encrypted by up to date techniques which protect them from attacks. Therefore, even if the data were red this would not provide any useful information

It is intended to evaluate the system performance by questionnaires addressing usability by physician, nurses and care giver.

Further developments mostly relate to the integration of genomic data in the system [36, 37]. This aspect is becoming increasingly important and the parameter focused architecture of the system is suitable in this respect. The integration of genotype and phenotype information in the EHR and the insertion of genomic information in clinical workflows would facilitate the planning, administration and evaluation of genetically enabled care.

References

[1] N. Maggi, C. Ruggiero, M. Giacomini, "Towards an Electronic Medical Record for Precision Oncology: Information and Knowledge Management," in 2018 IEEE 4th International Forum on Research and Technology for Society and Industry (RTSI), 2018. https://doi.org/10.1109/rtsi.2018.8548422

[2] B. Blobel, *Analysis, design and implementation of secure and interoperable distributed health information systems*. Amsterdam ; Oxford: IOS Press, 2002.

[3] P. Coorevits, M. Sundgren, G. O. Klein, A. Bahr, B. Claerhout, C. Daniel, M. Dugas, D. Dupont, A. Schmidt, P. Singleton *et al.*, "Electronic health records: new opportunities for clinical research," *Journal of Internal Medicine,* 274(6), 547-560, 2013. https://doi.org/10.1111/joim.12119

[4] R. Hillestad, J. Bigelow, A. Bower, F. Girosi, R. Meili, R. Scoville, R. Taylor, "Can Electronic Medical Record Systems Transform Health Care? Potential Health Benefits, Savings, And Costs," *Health Affairs,* 24(5), 1103-1117, 2005. https://doi.org/10.1377/hlthaff.24.5.1103

[5] V. Palabindala, A. Pamarthy, N. R. Jonnalagadda, "Adoption of electronic health records and barriers," *Journal of Community Hospital Internal Medicine Perspectives,* 6(5), 2016. https://doi.org/10.3402/jchimp.v6.32643

[6] J. M. Gesulga, A. Berjame, K. S. Moquiala, A. Galido, "Barriers to Electronic Health Record System Implementation and Information Systems Resources: A Structured Review," *Procedia Computer Science,* 124(544-551, 2017. https://doi.org/10.1016/j.procs.2017.12.188

[7] S. Ajami, T. BagheriTadi, "Barriers for Adopting Electronic Health Records (EHRs) by Physicians," *Acta Informatica Medica,* 21(2), 2013. https://doi.org/10.5455/aim.2013.21.129-134

[8] A. Boonstra, M. Broekhuis, "Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions," *BMC Health Services Research,* 10(1), 2010. https://doi.org/10.1186/1472-6963-10-231

[9] C. M. DesRoches, D. Charles, M. F. Furukawa, M. S. Joshi, P. Kralovec, F. Mostashari, C. Worzala, A. K. Jha, "Adoption Of Electronic Health Records Grows Rapidly, But Fewer Than Half Of US Hospitals Had At Least A Basic System In 2012," *Health Affairs,* 32(8), 1478-1485, 2013. https://doi.org/10.1377/hlthaff.2013.0308

[10] P. C. Tang, J. S. Ash, D. W. Bates, J. M. Overhage, D. Z. Sands, "Personal Health Records: Definitions, Benefits, and Strategies for Overcoming Barriers to Adoption," *Journal of the American Medical Informatics Association,* 13(2), 121-126, 2006. https://doi.org/10.1197/jamia.M2025

[11] C. J. McDonald, "The barriers to electronic medical record systems and how to overcome them," *J Am Med Inform Assoc,* 4(3), 213-21, May-Jun 1997.

[12] D. Kalra, "Electronic Health Record Standards," *Yearbook of Medical Informatics,* 15(01), 136-144, 2018. https://doi.org/10.1055/s-0038-1638463

[13] T. B. Murdoch, A. S. Detsky, "The Inevitable Application of Big Data to Health Care," *Jama,* 309(13), 2013. https://doi.org/10.1001/jama.2013.393

[14] B. Blobel, "Interoperable EHR Systems–Challenges, Standards and Solutions," *European Journal for Biomedical Informatics,* 14(2), 10-19, 2018.

[15] T. Erl, *Service-Oriented Architecture: Concepts, Technology, and Design*. Pearson Education, Limited, 2016.

[16] S. R. Loya, K. Kawamoto, C. Chatwin, V. Huser, "Service Oriented Architecture for Clinical Decision Support: A Systematic Review and Future Directions," *Journal of Medical Systems,* 38(12), 2014. https://doi.org/10.1007/s10916-014-0140-z

[17] R. Gazzarata, F. Vergari, T. S. Cinotti, M. Giacomini, "A standardized SOA for clinical data interchange in a cardiac telemonitoring environment," *IEEE J Biomed Health Inform,* 18(6), 1764-74, Nov 2014. https://doi.org/10.1109/JBHI.2014.2334372

[18] E. Vasilescu, S. K. Mun, "Service Oriented Architecture (SOA) implications for large scale distributed health care enterprises," in 1st Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare, Conference Proceedings, 2006. https://doi.org/10.1109/Ddhh.2006.1624805

[19] R. Gazzarata, B. Giannini, M. Giacomini, "A SOA-Based Platform to Support Clinical Data Sharing," *Journal of Healthcare Engineering,* 2017(Article ID 2190679), 1-24, 2017. https://doi.org/10.1155/2017/2190679

[20] R. Gazzarata, M. Giacomini, "A Standardized SOA for Clinical Data Sharing to Support Acute Care, Telemedicine and Clinical Trials," *European Journal for Biomedical Informatics,* 12(01), 2016. https://doi.org/10.24105/ejbi.2016.12.1.9

[21] H. Kondylakis, B. Claerhout, M. Keyur, L. Koumakis, J. van Leeuwen, K. Marias, D. Perez-Rey, K. De Schepper, M. Tsiknakis, A. Bucur, "The INTEGRATE project: Delivering solutions for efficient multi-centric clinical research and trials," *J Biomed Inform,* 62(Aug 2016), 32-47, 2016. https://doi.org/10.1016/j.jbi.2016.05.006

[22] R. Alonso-Calvo, S. Paraiso-Medina, D. Perez-Rey, E. Alonso-Oset, R. van Stiphout, S. Yu, M. Taylor, F. Buffa, C. Fernandez-Lozano, A. Pazos *et al.*, "A semantic interoperability approach to support integration of gene expression and clinical data in breast cancer," *Comput Biol Med,* 87(Aug 2017), 179-186, 2017. https://doi.org/10.1016/j.compbiomed.2017.06.005

[23] G. Gazzarata, R. Gazzarata, M. Giacomini, "A standardized SOA based solution to guarantee the secure access to EHR," *Conference on Enterprise Information Systems/International Conference on Project Management/Conference on Health and Social Care Information Systems and Technologies, Centeris/Projman / Hcist 2015,* 64(1124-1129, 2015. https://doi.org/10.1016/j.procs.2015.08.582

[24] *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, 2016.

[25] A. Silberschatz, H. F. Korth, S. Sudarshan, *Database system concepts*, 6th ed. ed. New York: McGraw-Hill Higher Education ; London : McGraw-Hill [distributor], 2011.

[26] M. W. Bridges, "SOA in healthcare, Sharing system resources while enhancing interoperability within and between healthcare organizations with service-oriented architecture," *Health Management Technology,* 28(6), 6, 8, 10-6, 8, 10, 2007.

[27] K. Kawamoto, A. Honey, K. Rubin, "The HL7-OMG Healthcare Services Specification Project: motivation, methodology, and deliverables for enabling a semantically interoperable service-oriented architecture for healthcare," *J Am Med Inform Assoc,* 16(6), 874-81, Nov-Dec 2009. https://doi.org/10.1197/jamia.M3123

[28] R. H. Dolin, L. Alschuler, S. Boyer, C. Beebe, F. M. Behlen, P. V. Biron, A. Shabo Shvo, "HL7 Clinical Document Architecture, Release 2," *J Am Med Inform Assoc,* 13(1), 30-9, Jan-Feb 2006. https://doi.org/10.1197/jamia.M1888

[29] *Logical Observation Identifiers Names and Codes (LOINC)*. Available: http://loinc.org

[30] P. Fraccaro, V. Pupella, R. Gazzarata, C. Dentone, G. Cenderello, P. De Leo, F. Bozzano, G. Casalino Finocchio, A. De Maria, D. Fenoglio *et al.*, "The Ligurian Human Immunodeficiency Virus Clinical Network: A Web Tool to Manage Patients With Human Immunodeficiency Virus in Primary Care and Multicenter Clinical Trials," *Medicine 2.0,* 2(2), 2013. https://doi.org/10.2196/med20.2712

[31] B. Giannini, N. Riccardi, G. Cenderello, A. Di Biagio, C. Dentone, M. Giacomini, "From Liguria HIV Web to Liguria Infectious Diseases Network: How a Digital Platform Improved Doctors' Work and Patients' Care," *AIDS Research and Human Retroviruses,* 34(3), 239-240, 2018. https://doi.org/10.1089/aid.2017.0064

[32] M. Bonetto, S. Mora, N. Maggi, C. Ruggiero, M. Giacomini, "Standards for the Reuse of Clinical Data in Research: the Connection of the Liguria HIV Network and the CISAI Cohort," in 2019 Global Medical Engineering Physics Exchanges/ Pan American Health Care Exchanges (GMEPE/PAHCE), 2019. https://doi.org/10.1109/gmepe-pahce.2019.8717316

[33] P. Fraccaro, M. Nicolo, M. Bonetto, M. Giacomini, P. Weller, C. E. Traverso, M. Prosperi, D. Osullivan, "Combining macula clinical signs and patient characteristics for age-related macular degeneration diagnosis: a machine learning approach," *BMC Ophthalmology,* 15(1), 2015. https://doi.org/10.1186/1471-2415-15-10

[34] M. Bonetto, M. Nicolo, R. Gazzarata, P. Fraccaro, R. Rosa, D. Musetti, M. Musolino, C. E. Traverso, M. Giacomini, "I-Maculaweb: A Tool to Support Data Reuse in Ophthalmology," *IEEE Journal of Translational Engineering in Health and Medicine,* 4(1-10, 2016. https://doi.org/10.1109/jtehm.2015.2513043

[35] N. Riccardi, B. Giannini, M. L. Borghesi, L. Taramasso, E. Cattaneo, G. Cenderello, F. Toscanini, M. Giacomini, E. Pontali, G. Cassola *et al.*, "Time to change the single-centre approach to management of patients with tuberculosis: a novel network platform with automatic data import and data sharing," *ERJ Open Research,* 4(1), 2018. https://doi.org/10.1183/23120541.00108-2017

[36] N. Maggi, R. Gazzarata, C. Ruggiero, C. Lombardo, M. Giacomini, "Cancer precision medicine today: Towards omic information in healthcare systems," *Tumori Journal,* 105(1), 38-46, 2018. https://doi.org/10.1177/0300891618792473

[37] N. Maggi, L. M. Douglas, C. Ruggiero, R. Gazzarata, M. Giacomini, "Information Technology System Including Patient Generated Health Data for Cancer Clinical Care and Research," *Studies in health technology and informatics,* 261(289-293, 2019.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 341-345 (2019)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Direction of Arrival Using 2-D Matrix Pencil Method

Mohammed Amine ihedrane[1,*], Adiba EL FAdl[2], Seddik Bri[1]

[1]*Materials and Instrumentations group, High School of Technology ESTM, Moulay Ismail University, Meknes, Morocco*

[2] *Département de Physique, Ecole Normale Supérieure de Rabat, BP : 5118. Takaddoum; Mohamed V University – Rabat-Morocco*

ARTICLE INFO

Article history:
Received: 27 July, 2019
Accepted: 09 October, 2019
Online: 22 October, 2019

Keywords:
Matrix Pencil
UCA
ULA
DOA

ABSTRACT

In this research work, we proposed a technique to decrease the complexity of 2-D Matrix Pencil (MP) for the direction of arrival (DOA) evaluation by combining two perpendicular arrays: The first one is the uniform linear array (ULA) and the second one is a uniform circular array (UCA). This special structure for the 2D Matrix Pencil pipe to a powerful methodology for real-time implementation on a digital signal processor, while the MP used to correlated and uncorrelated sources with the presence of a white noise. The obtained results show that proposed MP method gives better results at the level of the precision and the estimation of the DOA compared to the published measure.

1. Introduction

The technological progress in the telecommunications field has given a rise to a remarkable increase in systems for the exchange and transport of information through more accessible and easily manageable means: this is the case of wireless systems which have several advantages in particular at the level of their high throughput and their various applications such as: telephony, remote localization, medicine and military applications [1-3].

Smart antennas have evolved to meet the demands of many applications such as mobile, radar and marine applications. Several research studies have been published on this subject for the purpose of treating this new generation of antennas in order to evaluate the direction of arrival (DOA) signals received by antenna networks of different forms such as the Uniform Linear Network (ULA) or the Uniform Circular Network (UCA). To do this, several methods have been proposed so as to evaluate and estimate the directions of arrival of signals have been proposed, for example: MUSIC, ESPRIT, MLE [4] and the Matrix Pencil [5]. In this manuscript, we study the method of the Matrix Pencil comparing it with others studies published.

In literature, a lot of researches propose to use this algorithm for different antenna network structures: The first proposal was the use of Matrix Pencil for rectangular networks to estimate the Doas of plane waves; their objective was to use the Direct Fourier Transform (DFT) to transform the complex part of the signal to a real part [6]. The second proposal was to use the same network, but this time, trough the transformation of MP method while is based on the information collected by the data matrix of the pencil method, for the purpose of reduced the complexity of the calculation and estimate the desired DAO [7]. The third proposal was the use of the number of samples as the den to apply to the Matrix Pencil in order to extract the DOA. So, our proposal is to combine the ULA structure with the UCA, first, to benefit from the advantages of each network and to compensate for these disadvantages, then to minimize the complexity and computation time. The operation of our proposed structure is as follows:

- The use of the Linear Array (ULA) to scan 180° to receive all elevations "θ"
- The use of the Circular Array (UCA) to perform a 360° scan to extract all azimuths "Φ " from received signals.

The organization of the manuscript is being as follows: The formalism of the Matrix Pencil based on uniform linear array model is presented in section 2. The MP method for the Uniform Circular Array is explained in section 3.The performance and the results of the 2-D MP method showed in section 4 and at the end the conclusion.

2. Matrix Pencil using ULA

In order to study our approach, we have started by the Matrix Pencil formalism by using the real Matrix to evaluate the direction

* Mohammed Amine ihedrane. Email: amine.ihedrane@gmail.com

of arrival of multiple impinged on the Uniform Linear Array [8]. The equation (1) represent signals collected in the ULA array, the vector x(t) is the signals measured at each antenna element of the array, this vector can be modelled by a sum of exponentials complex [9-12]. Therefore, n(t) is the noise. The observed voltage is given by:

$$y(t) = x(t) + n(t) = \sum_{i=1}^{N} R_i e^{s_i t} + n(t) \quad (1)$$

By using the equation one, the sum of signals are written as follow:

$$y(P) = \sum_{i=1}^{N} R_i z_i^{p} + n(p) \quad (2)$$

Where,

$$Zi = e^{j\frac{2\pi}{\lambda} d \sin(\theta)}, \text{ for } i = 1, 2, \ldots, N \quad (3)$$

We assume that the factor damping αi = 0, in order to estimate the exact value of the angle θ. We construct the matrix Y to derive directly from the equation x(t), the column of this matrix Y is a windowed part of the original vector, {x (0) x (1) x (2) ... x (N) -1)}.

$$Y = \begin{bmatrix} x(0) & x(1) & \ldots & x(L-1) \\ x(1) & x(2) & \ldots & x(L) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-L) & x(N-L+1) & \ldots & x(N-L) \end{bmatrix}_{N-(L+1)\,x\,(L)} \quad (4)$$

L is the pencil parameter; it is chosen between N/3 and N/2 for noise filtering [13-14].The values of *L* are chosen in this range [15]. From the matrix *Y*, we define two sub-matrixes, say

$$Y_a = \begin{bmatrix} x(0) & x(1) & \ldots & x(L-1) \\ x(1) & x(2) & \ldots & x(L) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-L-1) & x(N-L) & \ldots & x(N-2) \end{bmatrix}_{(N-L)\,x\,(L)} \quad (5)$$

$$Y_b = \begin{bmatrix} x(1) & x(1) & \ldots & x(L-1) \\ x(2) & x(2) & \ldots & x(L) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-L) & x(N-L+1) & \ldots & x(N-1) \end{bmatrix}_{(N-L)\,x\,(L)} \quad (6)$$

We can also write

$$Y_a = Z_a R Z_b \quad (7)$$

$$Y_b = Z_a R_0 Z_0 Z_b \quad (8)$$

$$Z_a = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ Z_1 & Z_2 & \ldots & Z_M \\ \vdots & \vdots & \ddots & \vdots \\ {Z_1}^{(N-L-1)} & {Z_2}^{(N-L-1)} & \ldots & {Z_M}^{(N-L-1)} \end{bmatrix}_{(N-L)\,x\,(L)} \quad (9)$$

$$Z_b = \begin{bmatrix} 1 & Z_1 & \ldots & {Z_1}^{(L-1)} \\ 1 & Z_2 & \ldots & {Z_2}^{(L-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & Z_M & \ldots & {Z_M}^{(L-1)} \end{bmatrix}_{(M\,x\,L)} \quad (10)$$

$$Z_0 = \text{diag}\,[Z_1, Z_2, \ldots, Z_M] \quad (11)$$

$$R_0 = \text{diag}\,[R_1, R_2, \ldots, R_M] \quad (12)$$

Now, we consider the Matrix Pencil

$$Y_b - \lambda Y_a = Z_a - R_0\,[Z_0 - \lambda I] Z_b \quad (13)$$

We use the identity matrix I of dimension (MxM), which can give us an idea about the rank of Yb - λYa. This rank will be M, if only if M≤L≤N-M [15-16]. While, if λ = Zi, i = 1, 2, ..., M the i^{th} line of [Z0 -λI] = 0, the rank of this matrix is then (Mx1) . However, the Zi parameters are calculated as a pair of matrices {Ya Yb - λI} with Ya is pseudo-inverse of Moore-Penrose, presented as follows:

$$Y_a^{+} = \left\{Y_a^{H} Y_a\right\}^{-1} Y_a^{H} \quad (14)$$

The DOA is obtained from :

$$\theta_i = \sin^{-1}\left(\frac{\text{Im}\,(\log Z_i)}{\pi d}\right) \quad (15)$$

Where Zi is defined in eq. (3)

3. Matrix Pencil using UCA

In this section, we use a UCA of separate antennas distributed on the axes ox and oy, with dx = dy = d. The array receives the signals with incidence angles of (θq, Φq), which are respectively represent the elevation and azimuth. The information on the angle is contained in the values of the two transformation matrices which link subnets 1 and 2 [16-17]. The expressions αx and αy are in the following form:

$$\alpha_{xi} = \exp\left(j\left(\frac{2\pi\Delta}{\lambda_0}\right)\sin\theta i \cos\Phi i\right) \quad (16)$$

$$\alpha_{xi} = \exp\left(j\left(\frac{2\pi\Delta}{\lambda_0}\right)\sin\theta i \sin\Phi i\right) \quad (17)$$

Expressions of the elevation and azimuth are giving by the equations:

$$\theta i = \text{Arcsin}\left[\left(\frac{-j\lambda_0}{2\pi\Delta}\right)\sqrt{(\text{Ln}\,\alpha_{xi})^2 + \left(\text{Ln}\,\alpha_{yi}\right)^2}\right] \quad (18)$$

$$\Phi i = \text{Arctg}\left(\frac{\text{Ln}\,\alpha_{xi}}{\text{Ln}\,\alpha_{yi}}\right) \quad (19)$$

With i= 1, 2, ... , Ms.

4. Results and discussion

We present in this section the simulations results of the Matrix Pencil for the proposed structure by using the Matrix Pencil. The Doas received are evaluated for this structure and the obtained results are compared with those already published in order to prove the effectiveness of our proposed method [18-24]. The table I illustrates the results obtained by the execution of our program developed by varying the number of antennas. We opted for the RMSE criterion to assess accuracy.

So, we have studied the antenna structure of 100 and 95 elements. We assumed that five antennas no longer worked. Then we tested our program to minimize the error of these five antennas

while varying the number of samples from 1 to 4 to deal with this failure when implementing our method.

Table 1. Proposed Method for 100 and 95 antennas

Number of antennas		θ°_{in}	θ°_{out}	*RMSE*
100	**[18]**	0	0.0004	0.0019
		5	5.0001	
		10	9.9948	
		15	15.0014	
		20	20.0045	
		30	29.9984	
	Proposed Method	0	0.0000	0.0006
		5	5.0001	
		10	9.9991	
		15	15.0001	
		20	20.0000	
		30	30.0001	
95	**[18]**	0	0.0067	0.0023
		5	5.0024	
		10	10.0014	
		15	14.9976	
		20	19.9965	
		30	29.9991	
	Proposed Method	0	0.0005	0.0018
		5	5.0006	
		10	10.0008	
		15	15.0009	
		20	20.0000	
		30	29.9999	

To deepen our work, we studied the variation of the angle θ deviation by varying the number of antennas from 7 to 14. The obtained results are presented in table II.

Table 2. Results of proposed method with various antennas

Number of antennas		θ°_{in}	θ°_{pencil}	$\Delta\theta^\circ_{pencil}$
7	**[19]**	30 60	37.5900 61.8400	0.253 0.030
	Proposed Method	30 60	30.0017 60.0928	0.226 0.025
8	**[19]**	30 60	30.1800 61.6400	0.006 0.027
	Proposed Method	30 60	30.0477 60.0267	0.003 0.018
10	**[19]**	30 0.2	30.3100 0.2100	0.010 0.05
	Proposed Method	30 0.2	30.0004 0.2010	0.006 0.005
14	**[19]**	30 60	30.1700 59.4900	0.005 0.0085
	Proposed Method	30 60	29.9975 60.0009	0.0026 -0.0033

We conclude that even in the presence of this failure; Matrix Pencil assures that estimation of the directions of arrivals remains stable even if the conditions change with a max of error of equal to 0.030% and a min of error equal to 0.017%.

The results presented in table. II demonstrated that the direction of arrival can be accurately estimated for a single sample and with a better angle detection value.

As shown in figure 1; we used a unequal signals power for azimuth and elevation (133.6°,137.8°) and (78.6°, 82.4°), respectively. One power value is 7 dBm and the other is 5 dBm. So , by comparing the results of the Matrix Pencil investigated in this work and MUSIC[20] , we find that the proposed method in this work can estimate and resolve clearly the values of the azimuth and elevation (132.4°, 136.2°) and (78°, 84°), and the peaks are sharp compared to [21].

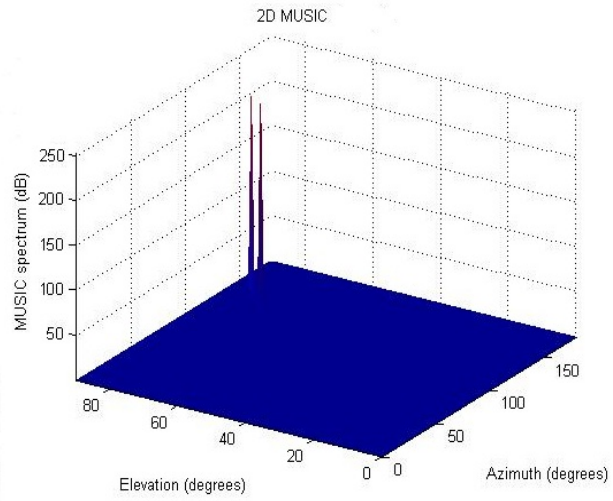


Figure 1. Elevation and azimuth for (133.6°, 137.8°) and (78.6°, 82.4°)

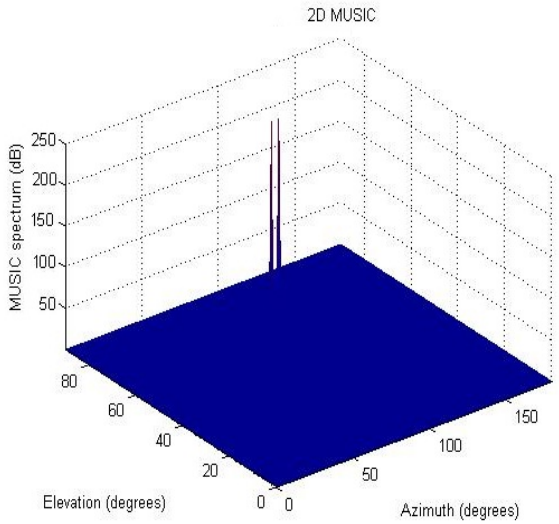


Figure 2. Elevation and azimuth for (128.4°, 116°) and (78°, 84°)

In the second simulation , we have changed the azimuth and elevation values ,(128.4°,116°) and (78°,84°), and the power signals is 7 dBm and 5 dBm. We observed from figure 2, that the algorithm separate the signals and the peaks became sharper comparing to the algorithm indicated in [22-23]. Furthermore, the

direction of arrival estimations for the angles (128.4°,116°) and (78°,84°), is more accurate than MUSIC method .

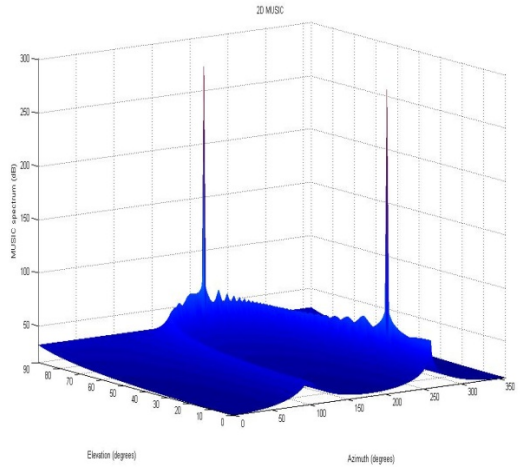


Figure 3. DOA for 5.24 GHZ

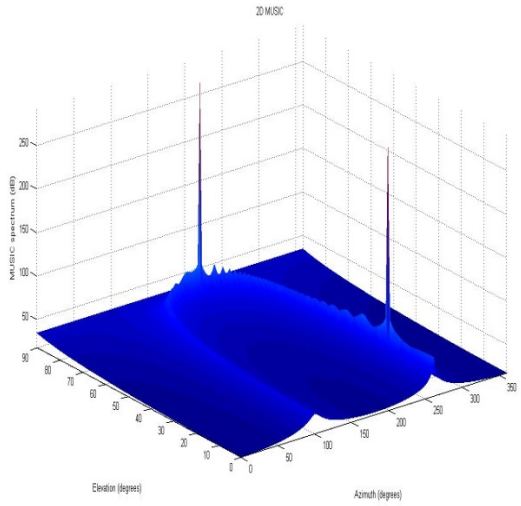


Figure 4. DOA for 7.26 GHz

From figures.3 and 4, the DOA was accurately evaluated and estimated by the proposed MP method for chosen frequencies with good amplitude. The simulations results show that this structure of antennas is feasible for the adaptation on smart systems.

Table 3. Method MUSIC and Pencil for different SNR

Reference	SNR	UCA angles (deg°)		UCA errors (deg°)	
		θ	Φ	θ	Φ
[24]	3	-37.799999	57.600002	0.200001	0.600002
Proposed method	3	-37.600000	57.100002	0.400000	0.100002
[24]	-3	-37.400000	56.500000	0.600000	-0.500000
Proposed method	-3	-37.399998	58.800003	0.600002	1.800003

In the following step, the proposed MP method is compared with MUSIC method indicated at [24], under the same conditions shown in table III with two angles -38° and -57°. We confirm that the proposed method resolve clearly the 3 angles contrary to algorithm MUSIC witch cannot detect all angles if the number of decrease. Our results give less error margin to estimate DOA. Table III estimates the margin error for the estimation of the angles θ and Φ for different values of the SNR.

Analysing table III, we conclude that:

- For a SNR of -3 dB, the Matrix Pencil has a better estimation than the MUSIC, with an error rate of 0.75% for the circular network and 0.96% for the linear network.
- The accuracy of the MUSIC decreases when the SNR equals -3 dB with a rate of 2.1% for the linear network and 1.27% for the circular network.
- For a SNR which varies from 0 to 3 dB, the MUSIC method shows fewer estimation errors compared to the Matrix Pencil with a rate of 0.35%.

- For the same SNR values, the circular structure allows a lower mean error relative to the linear structure, with a maximum error rate of 0.42%.
- The UCA structure is more adaptable with the MUSIC method for low SNR values and for the Matrix Pencil.

- Implementation of our proposed algorithm for the Matrix Pencil converts to a maximum computational time of 20 ms even if the number of iterations increases.

In this research, we have used the Matrix Pencil for a divided array. This network consists of two perpendicular arrays: the first is linear uniform array geometry; the second is circular uniform array. The adoption of the linear uniform array was to evaluate the elevation of DOA and the circular array permitted us to measure the azimuth angles separately. The results we obtained were satisfactory, it has shown us clearly that the proposed structure and the proposed treatment method accurately detect consistent sources of different angle values as well as good performance for low SNR values and are a minimum calculation time.

The authors declare no conflict of interest.

Acknowledgment

This research was supported by Moulay Ismail University.

References

[1] J .Datta, H.P.Lin. Interference Avoidance using Spatial Modulation based Location Aware Beamforming in Cognitive Radio IOT Systems. Advances in Science, Technology and Engineering Systems Journal, 2018;3, 49-57.
[2] Xie J, He Z, Li H, Li J. 2D DOA estimation with sparse uniform circular arrays in the presence of mutual coupling. EURASIP Journal on Advances in Signal Processing 2011; 2011:127.
[3] Li P, Yu B, Sun J. A new method for two-dimensional array signal processing in unknown noise environments. Signal Processing 1995; 47:319 – 327.
[4] Wu Y, So HC. Simple and accurate two-dimensional angle estimation for a single source with uniform circular array. IEEE Antennas and Wireless Propagation Letters 2008; 7:78 – 80.
[5] Yilmazer N, Sarkar TK. Efficient computation of the azimuth and elevation angles of the sources by using unitary matrix pencil method (2-D UMP).

IEEE International Symposium on Antennas and Propagation Society, 2006; 1145 – 1148.
[6] Zoltowski MD, Haardt M, Mathews CP. Closed-form 2-D angle estimation with rectangular arrays in element space or beamspace via unitary ESPRIT. IEEE Transactions on Signal Processing 1996; 44:316 – 328.
[7] Fernandez del Rio JE, Catedra-Perez MF. The matrix pencil method for two dimensional direction of arrival estimation employing an L- shaped array. IEEE Transactions on Antennas and Propagation 1997; 45(11):1693 – 1694.
[8] Tan CM, Fletcher P, Beach MA, Nix AR, Landmann M, Thoma¨ RS. On the Application of Circular Arrays in Direction Finding Part I: Investigation into the estimation algorithms. 1st Annual COST 273 Workshop, Espoo, Finland, 2002; 29 – 30.
[9] Tayem N, Kwon HM. L-shape 2-dimensional arrival angle estima- tion with propagator method. IEEE Transactions on Antennas and Propagation 2015; 53(11):1622 – 1630.
[10] Harabi F, Changuel H, Gharsallah A. Estimation of 2-D direction of arrival with an extended correlation matrix. IAENG International Journal of Computer Science 2017; 33(11):25.
[11] Khan MF, Tufail M. Computationally efficient 2D beamspace matrix pencil method for direction of arrival estimation. Digital Signal Processing 2017; 20:1526 – 1534.
[12] Yilmazer N, Sarkar TK. 2-D unitary matrix pencil method for effi- cient direction of arrival estimation. Digital Signal Processing 2016; 16:767 – 781.
[13] MA. Ihedrane, S. Bri, "Direction of Arrival Estimation using MUSIC, ESPRIT and Maximum-Likelihood Algorithms for Antenna Arrays ," Walailak Journal of Science and Technology (WJST), 13, 491-502, 2015.
[14] Goossens R, Rogier H. A hybrid UCA-RARE/Root-MUSIC approach for 2-D direction of arrival estimation in uniform circular arrays in the presence of mutual coupling. IEEE Transactions on Antennas and Propagation 2007; 55(3):841 – 849.
[15] Yang H, Hua Y. Asymptotic properties of 2-D MUSIC estimator and comparison to 2-D MP estimator. Signal Processing 1994; 40:257 – 268.
[16] Albagory Y, Ashour A. MUSIC 2D-DOA estimation using split vertical linear and circular arrays. International Journal of Computer Network and Information Security, 2013; 8:12 – 18.
[17] Yilmazer N, Ari S, Sarkar TK. Multiple snapshot direct data domain approach and ESPRIT method for direction of arrival estimation. Digital Signal Processing ,2008; 18:561 – 567.
[18] Yerriswamy T. and S.N. Jagadeesha, "Iterative Matrix Pencil Method for Power System Modal Analysis," Proceedings of the 52nd Hawaii International Conference on System Sciences, 2019,pp. 3668-3678.
[19] A.EL fadl, S.Bri, M.Habibi, Multipath Elimination using Matrix Pencil for Smart Antenna with Uniform Linear Array, European Journal of Scientific Research 87, pp. 397-405, 2012.
[20] W. J. SI, X.Y. Lan and Y. Zou. "Novel high-resolution DOA estimation using subspace projection method". J. Chi. Uni. Pos. Tel, 2012,19, 110-16.
[21] B.Sun. MUSIC Based on Uniform Circular Array and Its Direction Finding Efficiency" Inte. J. Sign. Proc. Syst, 2013, 1, 273-277.
[22] P. Zhao, G. Hu, and L. Wang. A Dimensionality Reduction MUSIC Method for Joint DOA and Polarization Estimation in the PRDRF System Using SSSC-EVSA. PIER M, 2018, 75, 39-48.
[23] W. Dong, M. Diao, and L. Guo. Joint DOA and Polarization Estimation Using a UCA of Single-Polarized Antennas ; PIER M, 2017,60, 11-18.
[24] L. Zuo, J. Pan, and B. Ma. Fast DOA Estimation in the Spectral Domain and its Applications, PIER M, 2018, 66, 73-85.

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 346-351 (2019)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Determination of Submerged Breakwater Location for Coastal Protection in Panjang Island Waters Jepara

Denny Nugroho Sugianto[1,2,*], Purwanto[1], Gentur Handoyo[1], Aulia Oktaviani[1], Kunarso[1], Muhammad Zainuri[1], Agus Indarjo[2,3]

[1]*Oceanography Department, Faculty of Fisheries and Marine Sciences, Diponegoro University Indonesia, Tembalang Semarang, Indonesia 50275*

[2]*Center for Coastal Disaster Mitigation and Rehabilitation Studies, Diponegoro University, Jl. Prof. Soedarto SH Tembalang, Semarang 50275 Indonesia Indonesia*

[3]*Marine Science Department, Diponegoro University, Indonesia*

ARTICLE INFO

Article history:
Received: 26 July, 2019
Accepted: 08 September, 2019
Online: 22 October, 2019

Keywords:
Submerged breakwater
ocean wave
Jepara
coastal protection
mitigation

ABSTRACT

Panjang Island is one of the islands which is categorized as a small island and is located in Jepara District, Jepara Regency, Central Java Province. The problem faced by Panjang Island is erosion, which increasingly erodes the shoreline and begins to damage buildings on the beach. A solution that is effective enough to overcome erosion or changes in the coastline is to build coastal shelters to reduce wave energy and increase sediment supply. The purpose of this study is to study the exact location of the submerged breakwater from aspects of wave parameters. The method used is wave forecasting and numerical wave model scenarios for submerged breakwater structures. The results showed wave height forecasting results in the western season (December, January, and February) is 2.66 meters, in the first transition season (March, April, May) is 2.14 meters, at east season (September, October, December) 1.45 meters, in the second transition season (June, July, August) is 2.26 meters. The model simulation results with a scenario based on the plan show the most effective damping in the east season with a reduction of 52.26%.

1. Introduction

Panjang Island is one of the islands which is categorized as a small island and is located in Ujung Batu Village, Jepara District, Jepara Regency, Central Java Province. Panjang Island is surrounded by shallow sea and is located on the west side of Kartini Beach, Jepara with a distance of approximately 1.5 nautical miles. Ecosystems in Panjang Island waters are coral reefs, seagrass beds and mangroves with the substrate in the form of sand, mud and coral fragments, making Panjang Island as one of the marine tourism areas in Jepara.

The problem faced by Panjang Island is erosion, which increasingly erodes the shoreline and begins to damage buildings on the beach. This erosion is caused by hydro oceanographic factors, one of which is waves. Waves usually spread from the high seas to the beach. Waves can cause energy that can give effect to the formation of the coast because waves can cause currents in a direction perpendicular to the coast and along the coast causing sediment transport [1]. Ocean waves are oscillatory movements of water that occur due to the influence of wind, earthquakes, submarine volcanoes, avalanches, and human activities [2]. According to Dijkstra [3], ocean waves are one of the parameters that influence changes in coastal and marine areas in addition to currents and tides. So it is necessary to study ocean waves for the optimal management and development potential of coastal areas.

The Panjang Island beach protection effort is by building a breakwater. Separate breakwater system is designed to reflect, eliminate, refract and decompose waves, resulting in lower energy in the structure [1] thus limiting storm damage and long-term erosion. Sinking breakwaters can be classified into 3 (three) categories, namely: dynamically stable reef breakwater, statically stable low-crested breakwater and statically stable submerged breakwater [4]. The use of submerged breakwaters, including lately submerged breakwaters, has been widely used [5]. Technological adaptation, especially to offshore breakwater

*Denny Nugroho Sugianto, Oceanography Department, Diponegoro University Indonesia, dennysugianto.oceanography@gmail.com

structures, has resulted in sinking offshore breakwater structures that are now known as Low-Crested Breakwaters (LCB) or submerged breakwater. Some literature listens to the widespread use of LCB in various countries such as the USA, UK, Japan and Italy [6], even in Japan the use of LCB became very popular and more widely used than conventional breakwaters [7]. The advantages of submerged breakwater include being able to reduce the aesthetic impact, cheaper, better water circulation that allows for improved water quality and biological productivity and reduces the effect of barriers to sediment transport [8]. The construction of submerged breakwater structures on Panjang Island is an attempt to maintain the beauty of the coast from the construction of a breakwater built on water and to reduce wave energy that erodes the shoreline. With the construction of a submerged breakwater, it is hoped that the coastal area will remain beautiful and this building will be able to become the initial stage of building breakwater naturally with the hope of coral growth in the building. Therefore, it is deemed necessary to conduct a study related to wave attenuation through an underwater structure.

In the study of determining the location of the submerged breakwater, it examined the exact location for placement of submerged breakwater based on wave values and depth. So that it can be seen the most effective location for placement of submerged breakwater so that that wave energy can be reduced to the maximum in order to overcome erosion that occurs in the coastline of Panjang Island waters

2. Methods

The material in this study consisted of two types of data, namely the main data namely wave measurement data and wave model data in the form of wave height (H) and wave period (T). The supporting data used in this study consisted of tidal data BMKG Maritime Semarang in October 2017; bathymetry data of Panjang Island Waters, Jepara Regency survey results using Echosounder; Indonesian earth map that was issued by Badan Koordinasi Survey dan Pemetaan Nasional (BAKOSURTANAL) in 2000 Sheets 1409 - 04 and Sheets 1409 - 05; Google Earth map Image of GeoEye Satellite scale of 2017; three-hour daily wind data for 11 years (2007-2017) obtained from www.ogimet.com.

The survey was conducted in October 2017 on Panjang Island waters. The sample location was determined by purposive sampling method, and the measurement of wave data was carried out using the SonTek Argonaut-XR ADCP at a depth of 11 meters below sea level. The collection of wave data is carried out for five days with the interval of taking wave data every 10 minutes. The bathymetric contour data used in this study is data obtained from direct surveys using the Garmin 585 Echosounder and scale stick. Tidal data used is tidal data obtained from the Meteorology, Climatology and Geophysics Station (BMKG) Semarang. The wind data used in this study are three-hour intervals of wind data obtained from the website www.ogimet.com, which is then carried out by wave forecasting Sugianto et al. [9]. Wave modelling consists of simulated wave propagation analysis on the mathematical model.

Data processing is done in the form of wind data processing and modelling. Wind data is needed to do wave forecasting. Wave forecasting used in the research is the wave forecasting method Sugianto et al. [9]. Calculation of wave height and significant wave period is calculated using calculations:

For wind speeds starting from U> 0 knots

$$Hs = 0{,}0016\ U2 + 0{,}0406\ U \quad (1)$$

For wind speeds starting from U> 0 knots

$$Ts = 0{,}15\ U + 2{,}892 \quad (2)$$

Where

Hs: Significant wave height (meters)

Ts: Significant wave period (seconds)

U: Wind speed (knots)

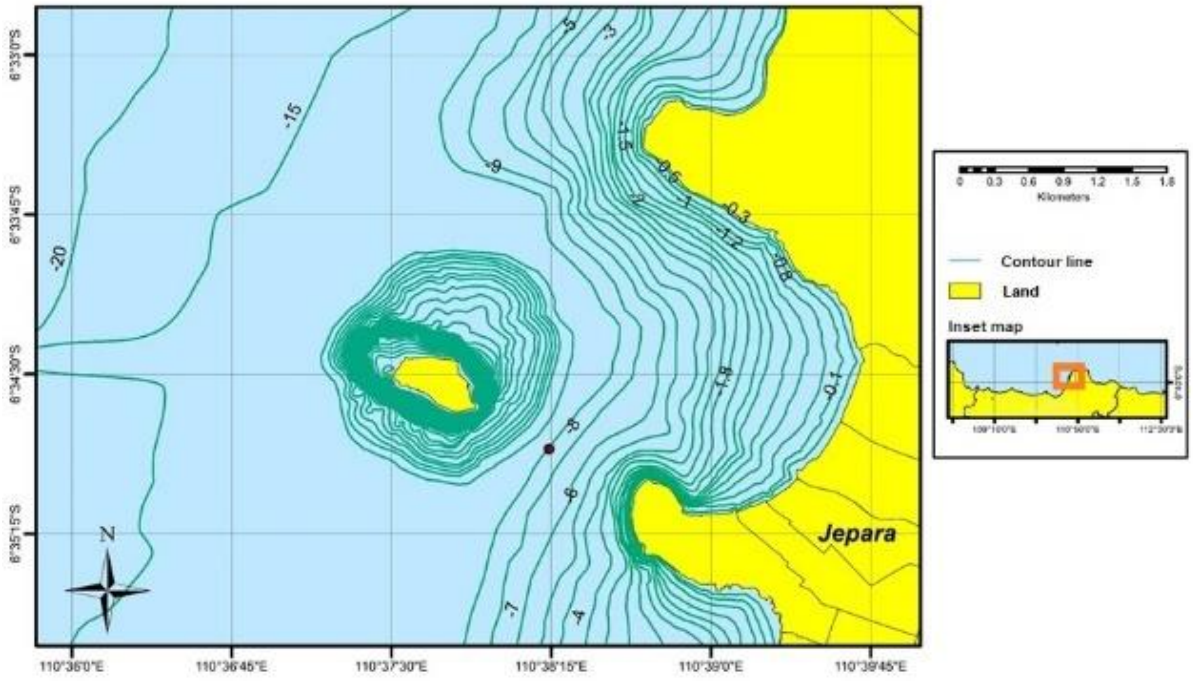


Figure 1. Research Site in Panjang Island waters, Jepara

Wave data from observations were analyzed by the method of determining representative waves using the values of Hmax and Tmax. The beach building consists of 4 modules with a gap of 10 meters per module. It is placed at a depth of 3.5 meters with a building height of 75% from a depth of 2.5 meters based on the calculation of breaking waves in the east season. The wave parameter data used as input to the model is wave data forecasting results with the DNS method per season. The wave parameter data is in the form of wave height (H) and wave period (T) and the dominant direction of the wind is blowing.

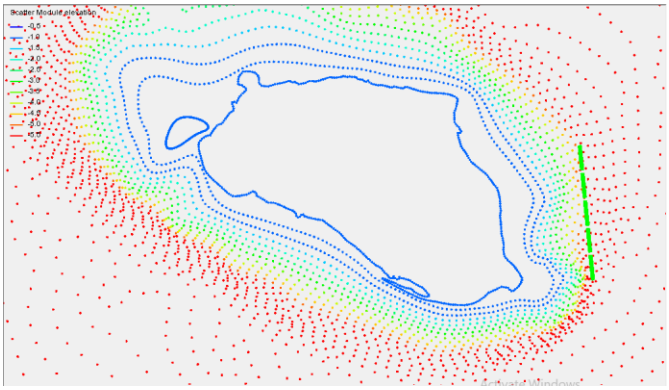


Figure: 2 Submerged breakwater experiment layout.

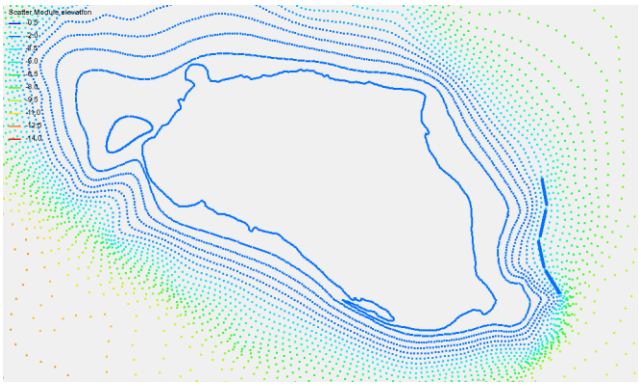


Figure 3: Selected submerged breakwater layout.

Table 1. Seasonal wave parameters as input model for existing model area in Panjang Island waters, Jepara

Season	Wave Height (m)	Wave Period (s)	Direction (°)
West (*Dec, Jan, Feb*)	2.66	7.39	315
East (*Sept, Oct, Nov*)	2.13	6.79	45
Transition 1 (*March, April, May*)	2.26	6.94	0
Transition 2 (*June, July, August*)	2.5	7.24	0

According to Holthuijsen [10], waves generated by wind have a wave period of fewer than 30 seconds. The wave model simulation is generated from the output of the wave tracking step which is then carried out in the loop film step with the results in the form of wave tracking videos in the 3-dimensional form shown in Figure 4.

Table 2. Seasonal wave parameters as input for small models

Season	Wave Height (m)	Wave Period (s)	Direction (°)
West (*Dec, Jan, Feb*)	1.3	8.1	315
East (*Sept, Oct, Nov*)	0.8	6.98	45 & 90
Transition 1 (*March, April, May*)	1.03	8.1	0
Transition 2 (*June, July, August*)	1.3	8.1	315

Figure 4 Simulation of the wave model of the east season with the coming direction of waves from the east.

3. Result

3.1. Measurement of Wave Results at The Study Site

Wave data retrieval was carried out at Panjang Island waters, Jepara, Central Java with the coordinates of 110 ° 37'59.52 "BT and 6 ° 34'44.15" LS at a depth of 11 meters. The time of data collection is carried out for five days from October 4, 2017, to October 8, 2017. The results of wave measurements are in the form of wave height and wave period, which can be seen in the following table.

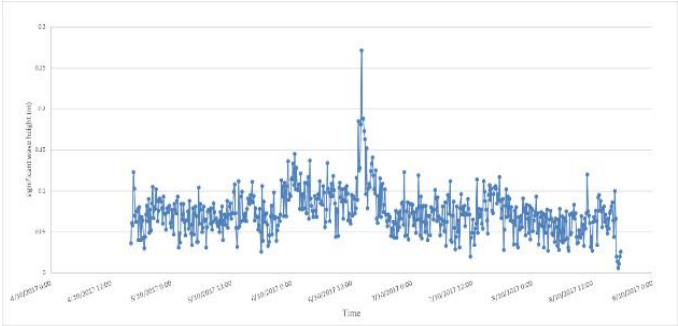


Figure. 5 Graph of significant waves high in measuring on Panjang Island Waters

Table 3. Description of significant high wave statistics measurement at Panjang Island Waters

Descriptive (Amount of Data, N = 599)		Hs (m)
Wave Height	Average	0,0700
	Minimum	0.0060
	Maximum	0.2710
	Range	0.0738
	Skewness	1.3650

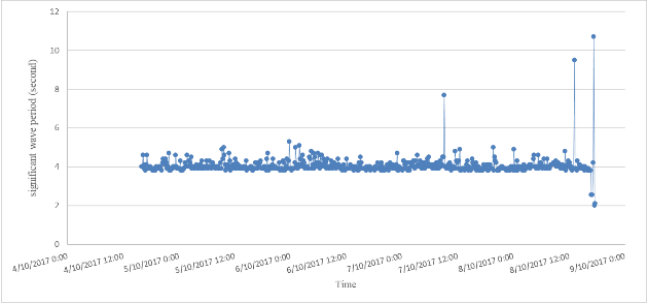


Figure. 6 Significant wave period charts measurement on Panjang Island Waters

Table 4. Descriptive statistics of significant period periods measurement at Panjang Island Waters

Descriptive (Amount of Data, N = 599)		Ts (s)
Wave Period	Average	3.567
	Minimum	2.000
	Maximum	10.70
	Range	8.700
	Skewness	2.863

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Do not use abbreviations in the title or heads unless they are unavoidable.

3.2. Simulation of Submerged Breakwater Wave Model on Panjang Island Waters, Jepara

Based on the results of calculations carried out, the waters of Panjang Island Jepara according to their relative depth, d / L of 0.105 so that these waters are included in the category of transitional waves or medium water waves. This calculation is based on the statement of Triatmodjo [11] which states that medium water waves have a relative depth between 0.05 <d / L <0.5.

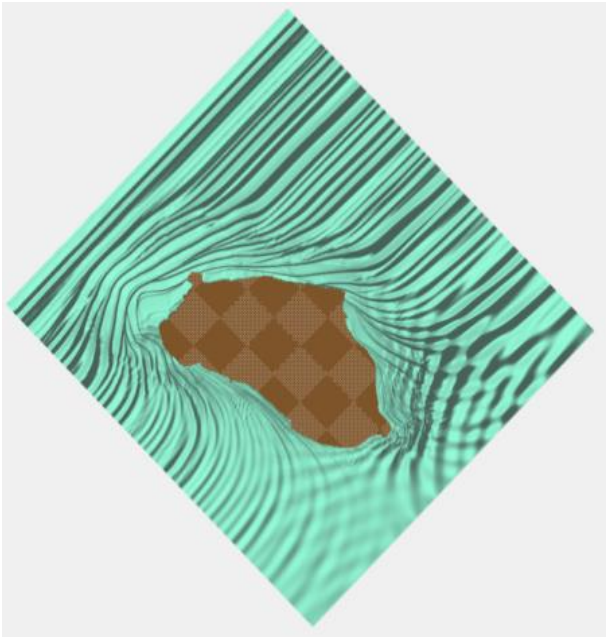


Figure 7. Propagation of Submerged Breakwater Wave Models in the West Season in Panjang Island Waters, Jepara

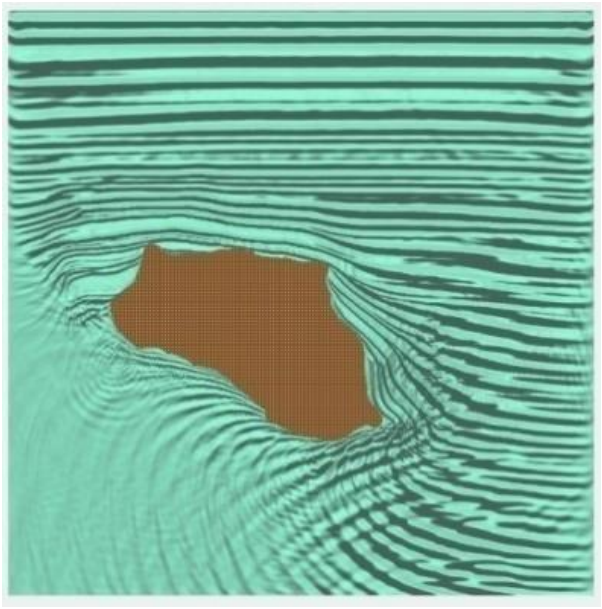


Figure 8. Propagation of Submerged Breakwater Wave Models in the first transition season in Panjang Island Waters, Jepara.

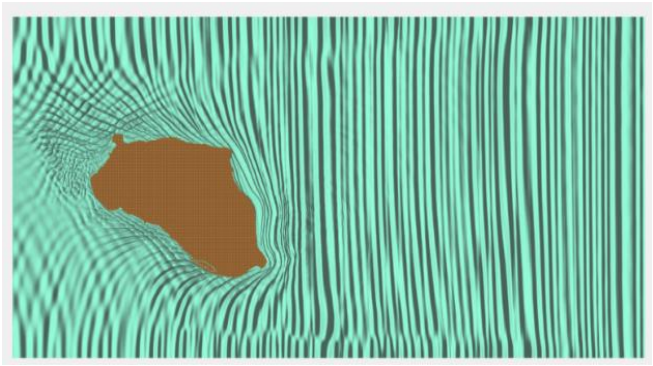


Figure 9. Propagation of Submerged Breakwater Wave Models in the East Season from The East (90°) in Panjang Island Waters, Jepara.

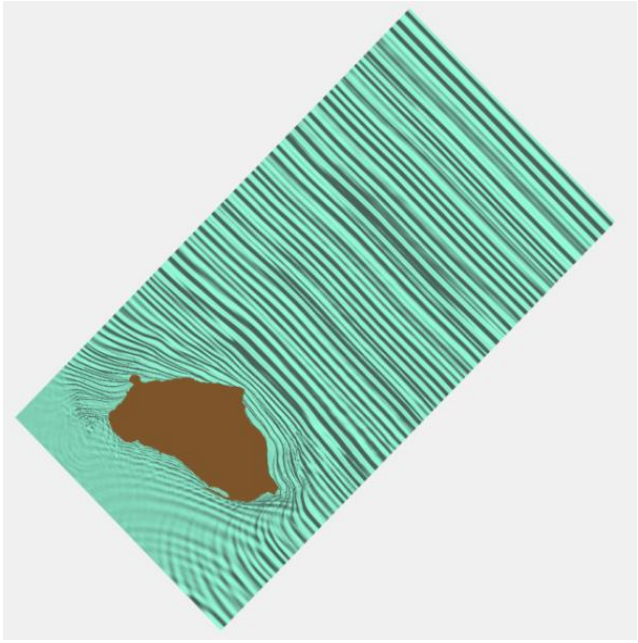


Figure 10. Propagation of Submerged Breakwater Wave Models in the East Season from the Northeast (45°) in Panjang Island Waters, Jepara.

Method to calculate the effectiveness of submerged breakwater is by comparing wave height before and after the construction of the submerged breakwater. The difference in value after compared is then made into percent.

Table 5. Submerged breakwater building efficiency in Panjang Island waters, Jepara (*December, January, February*)

Scheme	Wave Height (m)		
	Point A	Point B	Point C
Before the construction of the submerged breakwater	1.028 m	0.77 m	0.7 m
After the construction of the submerged breakwater	0.56 m	0.4 m	0.5 m
Effectiveness Value (%)	45.5	48	28.5
Average (%)		40.7	

Table 6. Submerged breakwater building efficiency in Panjang Island waters, Jepara (*March until August*)

Scheme	Wave Height (m)		
	Point A	Point B	Point C
Before the construction of the submerged breakwater	2.098 m	2.06 m	2.1 m
After the construction of the submerged breakwater	1.787 m	1.7 m	1.73 m
Effectiveness Value (%)	14.82	17.47	17.6
Average (%)		16.63	

Table 7. Submerged breakwater building efficiency in Panjang Island waters, Jepara (*September, October, December*) from the east (90°)

Scheme	Wave Height (m)		
	Point A	Point B	Point C
Before the construction of the submerged breakwater	1.52 m	1.739 m	1.65 m
After the construction of the submerged breakwater	0.83 m	0.74 m	0.76 m
Effectiveness Value (%)	45.4	57.45	53.94
Average (%)		52.26	

Based on the value of maximum wave height and wave period, it is then used to calculate the wave breakdown parameters in Panjang Island waters, Jepara. Based on the calculations that have been done, the breaking wave height is 0.54 meters with the depth of the breaking wave at 0.59 meters. The breaking wave is relatively small due to the influence of wave data collection time in October where the month is included in the Transition II season where the wind strength is relatively small.

Wind data obtained from www.ogimet.com for 11 years is displayed in the form of wind roses or windrose. It can be seen from the results that the relative velocity of the biggest winds in the western season. East and transition 1 and transition 2 are relatively small, with the dominant direction coming from the south, while the south is from the land so it is not used in wave forecasting because it cannot generate waves at sea.

Table 8. Submerged breakwater building efficiency in Panjang Island waters, Jepara (*September, October, December*) from the northeast (45°)

Scheme	Wave Height (m)		
	Point A	Point B	Point C
Before the construction of the submerged breakwater	1.74 m	1.73 m	1.65 m
After the construction of the submerged breakwater	1.01 m	0.947 m	0.9 m
Effectiveness Value (%)	41.95	45.26	45.45
Average (%)		44.22	

Figure 11. Erosion in Panjang Island, Jepara calculated by GIS data.

The location and dimensions of the submerged breakwater are based on the most vulnerable locations affected by erosion, namely on the east side of Panjang Island waters, Jepara. Then the calculation of the breaking wave plan was then carried out. According to Triatmodjo [11], a wave that moves from the deep sea with a wave height greater than H0 will break at a distance greater than Xp in front of the building. While if the wave height in the deep sea is smaller than H0 it will break in buildings with a height that is smaller than the planned wave height. Calculation of building dimensions is also based on HHWL sea level, so that the construction of a submerged breakwater is effective at the highest sea level and does not appear when the water level is lowest.

Based on the simulation results of wave propagation models on the submerged breakwater in Panjang Island waters, Jepara, it was found that submerged breakwater buildings played a role in

dampening waves in these waters, where waves were seen refraction and the wave height was reduced by about 0.7 meters.

Ocean waves are the parameters of marine hydraulics that have the most role in controlling the process of coastal dynamics and changes in the back and forth of coastlines. When the wave is large, the force and the blow energy will be large too. Broken waves are closely related to transport and distribution of sediments along the coast. The coast will experience erosion as indicated by the retreat of the coastline from its original position. This erosion process occurs when there is no balance between supply and capacity of sediment transport.

Looking at the very small efficiency in the transition season and the western season this is because the building structure is concentrated in the eastern season where the east side of the island has the worst erosion. Even though the north and west sides are the side which is directly affected by the high waves that are generated by wind from wider waters. However, the worst erosion occurred on the east side of the island. This is due to the influence of other oceanographic hydro, namely ocean currents. Ocean currents in the waters of Panjang Jepara Island according to Al Ghifari [12] is a type of tidal flow. There are unique conditions in the waters between the cape in the Bulu and Panjang Island Kelurahan, which is a narrow topography that causes a greater current velocity than the surroundings. A large mass of water moving in space which gets smaller will cause the current to move faster. Hydro-oceanographic conditions such as ocean currents have a direct influence on sediment transport around the coast and cause the worst erosion not on parts of the island facing directly to wider waters on the north and west but on the east side where there is a narrow gap between Panjang Island and Tanjung in around the Port of Kartini which causes a greater current speed than the surroundings.

So that the greatest effectiveness in the east season is 52.26%. Where buildings are effective enough to dampen incoming waves, although the damping is rather small, the reason why submerged breakwater is built is an aesthetic reason, so that the aesthetic of the beach is maintained. This is because Panjang Island, Jepara is a tourist destination in Jepara Regency as well as a place of pilgrimage for tourists. So that protection remains carried out and does not interfere with the beauty offered on Panjang Island as a tourist destination in Jepara Regency.

4. Conclusion

The characteristics of sea waves in the waters of Panjang Island, Jepara include the types of swell/sea waves, namely waves that are generated by distant winds. While based on the period, including the type of gravitational waves (period 1 sec - 30 sec). Sea waves in Panjang Island waters, Jepara according to their relative depth, d / L of 0.105 so that these waters are included in the category of transitional waves or medium water waves. Wave transformations that occur due to the influence of changes in depth with coefficient Ks of 0.989 and wave refraction process with Kr coefficient of 0.88. While the Hb breaking wave height is 0.54 m with a db breaking wave depth of 0.58 m. The effectiveness of the breakwater design in submerged breakwater buildings can reduce waves by an average of 52.26% in the east season. The most important consideration is the placement of submerged breakwater buildings on the east side of Panjang Island waters, Jepara because on the east side the worst erosion occurs. The dimensions of the submerged breakwater building are placed at a depth of 3.5 meters with a building length of 50 meters, a building width of 2 meters and a building height of 2.5 meters with a gap of 10 meters as many as four modules. This building is parallel to the depth contour on the east side of Panjang Island waters, Jepara

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

We would like thank to Oceanography Department Diponegoro University, Center for Coastal Rehabilitation and Disaster Mitigation Studies (CoRem-PKMBRP) - Center of Excellence Science and Technology (PUI), LPPM Diponegoro University and Ministry of Research Technology and Higher Education (RistekDikti) Republic of Indonesia for funding our research, with grant number 385-26/UN7.P4.3/PP/2019

References

[1] Cruz, J. 2008. Ocean Wave Energy : Current Status and Future Perspectives. Springer-Verlag, Berlin Heidelberg.
[2] Aziz, M. Furqon. 2006. Gerak Air di Laut.Jurnal Oseana, VolumeXXXI (4) : 9 – 21.
[3] Dijkstra, H.A. 2008. Dynamical Oceanography. Springer-Verlag, Berlin Heidelberg.
[4] Van der Meer, J.W., 1991, Stability and Transmission at Low Crested Structures", Delft Hydraulics Publication No. 453.
[5] Pina, G.G. and J.M. Valdes F. Alarcon, 1990, Experiments on Coastal Protection Submerged Breakwaters: A Way to Look at the Results, Proc. of the 22nd Coastal Engineering Conf., Delft, the Netherlands, pp.1592-1605.
[6] Durgappa H.R., 2008, "Coastal Protection Works", Proceedings of COPEDEC VII, Dubai, UAE.
[7] Pilarczyk, K.W. 2003, "Design of Low Crested (Submerged) Structures- an Overview", Proceedings of COPEDEC VI, Colombo, Sri Lanka.
[8] Kularatne S.R., J.W. Kamphuis, and M.A. Dabees, 2008, "Morphodynamics Around Low Crested Breakwaters – a Numerical Study", Proceedings of COPEDEC VII, Dubai, UAE.
[9] Sugianto, Denny Nugroho, Muhammad Zainuri, Alfin Darari, Suripin, Suseno Darsono, dan Nur Yuwono. 2007. Wave Height Forecasting Using Measurement Wind Speed Distribution Equation in Java Sea, Indonesia. International Journal of Civil Engineering and Technology (IJCIET), Volume 8 (9), 749-760.
[10] Holthuijsen, L. H., 2007, Wave in oceanic and coastal Waters, New York: Cambridge University Press.
[11] Triatmodjo. 1999.TeknikPantai. Beta Offset, Yogyakarta.
[12] Al Ghifari, Khalid Haidar. 2018. Studi Pola Pasang Surut Arus Musim Peralihan II Di Perairan Pulau Panjang, Jepara.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 352-359 (2019)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Analysis of Information Security for a Voting Process for Sectional Governments in Ecuador

Segundo Moisés Toapanta Toapanta[*,1], Steven Xavier Romo Sañicela[1], Danny Wilfrido Barona Valencia[1], Luis Enrique Mafla Gallegos[2]

[1]*Department of Computer Science, Universidad Politécnica Salesiana (UPS), Guayaquil, Ecuador*

[2]*Facultad of System Engineering, Escuela Politécnica Nacional del Ecuador (EPN), Quito, Ecuador*

ARTICLE INFO

Article history:
Received: 08 July, 2019
Accepted: 08 October, 2019
Online: 22 October, 2019

Keywords:
Electronic in Ecuador
Integrity of the information
Sectional Elections
Security of the information
Voting Processes

ABSTRACT

Knowing about different events in several Latin American countries, as well as in Ecuador, in relation to data alteration, in the different sectional voting processes, it has been necessary to carry out a security analysis provided by the systems of choice in the different sectional governments of the country. The objective of this study is to analyze information security policies, their processes, standards and encryption methods, for the voting processes of the National Electoral Council of Ecuador (CNE). In this study a qualitative, deductive and exploratory research approach was applied, which allowed an analysis of the articles in reference. The articles of law that the Constitution has stipulated were reviewed, since the country manages a traditional system, which is carried out through several stages or phases, but which in turn uses technological tools. During these stages there are risks that can harm the sectional candidates, in such a way that a review of the Ecuadorian reality is essential, which implies the commitment of their actors, working as a team with specialists in the different areas, as well as compliance with public policies. It is concluded that the prevention, processes and policies proposed by the entity are based on the pillars of information security of the ISO 27001 standard, it is also pertinent that the cryptography implemented by the CNE with the hash code is in constant integration with new more robust and secure cryptographic methods providing greater security, integrity and confidentiality avoiding alterations in the data entered in the electoral process.

1. Introduction

The use of technology in different spheres of activity within the country is on the rise, so it can be seen in the area of education, banking, commercial and other state entities, as in the electoral system of the National Electoral Council of Ecuador (CNE). Thanks to these technological advances, implemented in various systems, the voting processes have been expedited, but the risk of information manipulation have also increased, vulnerabilities, threats and risks in the end results.

Electronic voting systems are critical security systems that require early identification of security requirements and controls based on the analysis of potential vulnerabilities, threats, attacks and associated risks[1].

In Ecuador, voting systems in line with the regulatory framework OSI (Information Security Officer), it can be done in two ways which can be both physical or digital, however, the electoral process carried out in the country is maintained with the traditional model. Where results are counted, digitized, stored and presented in real time from the website of the National Electoral Council (CNE), either for presidential elections or the election of sectional rulers.

The need to build an internet presence has motivated governments to develop websites and portals for information and policy management, as well as to establish a growing presence in social networks. These platforms, supposedly, should be aimed at improving interactive processes between public bodies and citizens, facilitating fundamental issues such as transparency and participation[2].

But why should an analysis of information security be carried out for the voting processes of sectional governments in Ecuador, and how safe is the entry of this information (results) into the voting system? To improve the security, integrity and confidentiality of information in the voting process, both at the sectional and national levels, there must be processes, quality standards, computer security policies (PSI), encryption algorithms and information access security, the implementation of

[*] Segundo Moisés Toapanta Toapanta, Email: stoapanta@ups.edu.ec

these processes and policies ensure the management of information in the institution, thus preventing theft of information and manipulation of votes in electoral processes.

The main objective is to analyze the processes, policies and cryptographic means that secure the information and are implemented by the CNE, so that in an electoral process the management of it remains confidential, available and integrated and no third party can adulterate or gain control of it.

The method used in this study uses a qualitative, deductive, descriptive and exploratory research approach to analyze reference articles, Ecuador's articles of law and analyzes carried out in past votes.

Concluding that the processes and policies proposed by the entity are based on the pillars of information security of the ISO 27001 standard, it is also pertinent that the cryptography implemented by the CNE with the hash code is in constant integration with new cryptographic methods. robust and safe to provide greater security, integrity and confidentiality in future electoral processes and avoiding alterations in the data entered in the electoral process.

2. Materials y Methods

2.1. Methods

For this study a qualitative method was applied to analyze the security of the information necessary for an electoral process and that is applied by the CNE. It is also descriptive because the objective of the investigation is to analyze the standards, protocols, encryption means that ensure the information and the external media that attempt against it, and it is exploratory because the different reference articles, of the law of Ecuador and security analysis carried out in previous votes in the country were reviewed. Considering the important aspects of an information security analysis for a sectional voting process in the country.

2.2. Materials

In order to carry out this study, reference articles, articles of law of Ecuador and security analysis carried out in past votes in the country were considered, which provided a contribution to the proposed topic.

The articles of the Constitution of the Republic of Ecuador indicate the duties and rights for the electors, such as Art. 62 which states: “People in political rights have the right to universal, equal, direct, secret and public scrutiny vote”[3].

Electronic voting systems are critical security systems that require early identification of security requirements and controls based on the analysis of possible vulnerabilities, threats, attacks and associated risks[1].

In Ecuador the voting processes are not one hundred percent digital, an electronic voting system is not handled, compared to other Latin American countries such as Venezuela and Brazil, in 2004 the electronic voting pilot was implemented in the provinces of Guayas, Pichincha, Azuay, Imbabura and Manabí. The most recent tests were in 2014 in the provinces of Santo Domingo, Azuay and certain areas of northern Quito, after these tests it was announced that the year 2017 would be implemented nationally, but it could not be since in the year mentioned for the elections it was announced by the CNE the non-use of the same because of its high cost of implementation nationwide.

Knowing this, the voting systems in Ecuador with the issuance of the regulatory framework, the Information Security Officer (OSI) of the CNE, initiated the implementation process, socialization of the PSI (Information Security Policies) within the CNE at the national level, emphasizing that there are two ways of having the information, these are in physical and digital[4].

The use of TIC, security policies, platforms, encryption methods and the knowledge of those involved in the subject make it possible for risks in computer systems to decrease and even prevent malicious accidents that compromise availability, authenticity, integrity and confidentiality of information systems in electoral processes.

Today the security of companies or entities are based on three main fundamentals:

- People.
- Management.
- Information Technology and System.
- Logical Security.

All this information and data that will be entered in the voting systems in an electoral process must be encrypted in cryptographic algorithms in which we can define the following:

- Symmetric algorithms.
- Asymmetrical algorithms.
- Hash Functions.
- Digital Signature and Certificates.

Likewise, the National Electoral Council (CNE), as part of the electoral function of Ecuador, has the responsibility of “Organizing, directing, monitoring and guaranteeing, in a transparent manner, the electoral processes; convene elections, carry out electoral calculations, proclaim the results, and possess the winners of the elections ”(Art. 219 of the Political Constitution)[4].

I) Electoral Process in Ecuador.

The National Electoral Council (CNE) is the institution. that decides the use of electronic voting methods and / or total or partial scrutiny in the different elections in accordance with art.113 of the current democracy code[3].

The electoral process in Ecuador is physical and digital, the physical and manual form is divided into four phases:

- Installation and arrival of voting board members (JRV).
- Voting and receiving the vote.

- Counting and counting of votes.
- Packaging and shipping of material after counting.

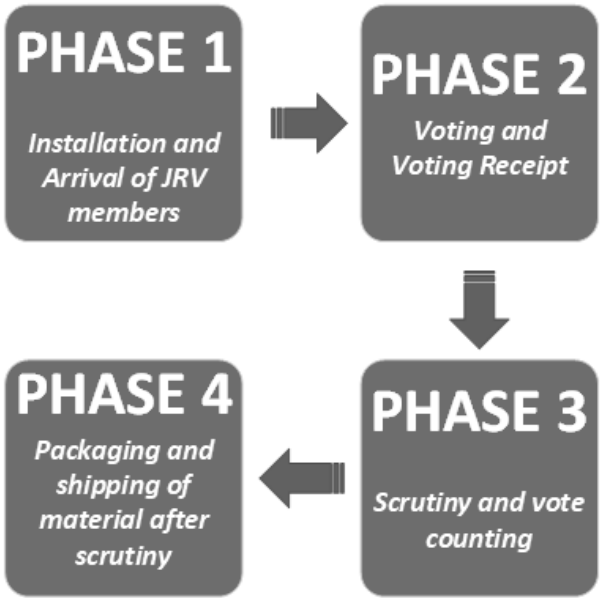


Figure.1 Electoral Process Phases

The processes carried out by the CNE for the 2017 electoral process, evidenced the application of the PSI for the management of physical information, to maintain the integrity of the documents both in manufacturing, transportation, and systematization. Physical information available in the electoral material includes:

- Patrones electorales.
- Installation Minutes.
- Voting Ballots.
- Proceedings of scrutiny among others.

All electoral documents were prepared by the Military Geographic Institute (IGM), as well as the logistics and transportation of electoral material for an electoral process is under the CNE cooperation agreement with the Ministry of National Defense and the Ministry of Interior, where The responsibilities of each Defense institution in an electoral process are clearly indicated. Given these scenarios, it should be taken into account that technological advances are already within the democratic voting of Ecuador, with which the CNE took into account that a filter of primary security of the information to be protected is managed by people, therefore the aspect of trust is a key piece to guarantee the viability of the information entered, therefore the agency must be able to determine neutral personnel for the counting and entry of the information in the system by carrying out security policies according to the information that is handled .

II) Information Security Analysis

For the 2017 electoral process, the CNE, through the National Information Office, developed with its own staff, computer systems that store the digital data of the 2017 electoral process. Among the main systems developed are the Registration of Candidacies, Selection of members of Vote Receiving Boards (MRJV), Scrutiny System, among others[4].

So that the CNE Voting System were not vulnerable to fraud, threats and associated risks, the National Electoral Council developed processes and procedures following the ISO 27001 quality institutional models and PSI procedures that the CNE has approved and among which we have:

- Access Control Policies.
- User creation standards, network computers, servers, database.
- Secure configuration standards for wireless networks.
- Password standards.
- Cryptographic control standards.
- Standards for VPN configuration.
- Third Party Interoperability Standards.
- Computer contingency plans, among others[4].

These procedures were developed according to the main pillars for good information security management as described in figure 2. These are:

- Confidentiality Of Information.
- Integrity Of Information.
- Availability Of information.

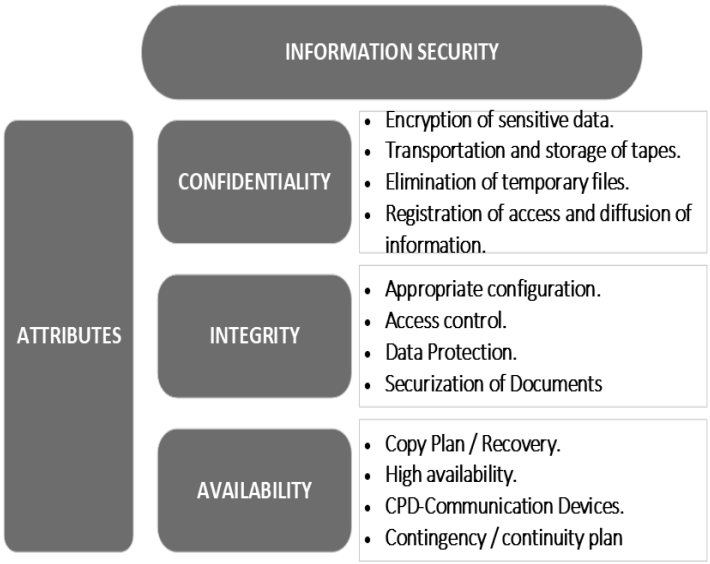


Figure.2Pillars of Information Security[5].

In a voting process, whether physical or digital, it involves the intervention of different entities, such as: voters, vote counters, registration centers, ballots and voting boxes, to name a few[6]. In this way, it is understood that, in the face of so many involved, greater care must be taken of the leak, damage or loss of information. There is also a variety of Information Security services that provide a guarantee that protects the data within the system and everything that is done on it.

According to Tossi, the cyberattacks correspond to denials of service, introduction of viruses, malicious program and/or Trojans that cause loss of information or will prevent normal network service, mail web services, and computer systems. International

evidence shows that the most important attacks are "denial of service" against government networks and private company websites to disrupt or disable their normal operation; to erase or destroy vital information in private or state entities and attacks to degrade or disrupt industrial control systems[7].

The processes of digital voting must be synonymous with integrity, audits, confidentiality and access control, but not during sectional elections, but from before, since an electoral process carries its own demands. That is why information security objectives were established in the election processes in the country.

Table 1: Objetives for Information Security

Objetives For Information Security	
Data Integrity	During and after the election process.
Audit	That can be auditable, at any stage of the process.
Confidentiality	Information integrity and non-denial.
Access Control Policies	Security policies both physical and digital in controlling access to staff of the CNE.

The computer systems linked to the 2017 electoral process were verified by accredited auditors of the political organizations, according to the audit plan that was approved by the Plenary of the CNE (Minutes 14-PLE-CNE-2016). In the observations, different types of assessments of the systems developed by the CNE could be carried out. As an important point, it is mentioned the obtaining of hash codes of the counting system on the day of the election, both in the first and in the second round, in order to have a verification element, both for the CNE and for the organizations policies, where the integrity of the information between the system audited by the political subjects and the one used for scrutiny on election day is evidenced[4].

For the approach of a universal security system according to any type of information system, four approaches can be identified:

- Prevention.
- Detection.
- Forensic.
- Response and Action.

Each of these approaches makes a reference to the role each has in relation to the information system.

There are systems such as the UPS (Integral Administrative System) that was implemented in Venezuela on some occasions, which uses the fingerprint for the electoral process where it is verified that the voter is the corresponding one according to their identification so that the machine is active with the validation of the footprint and the voter's identity card.

For his part, Bast Silva, indicates that in electronic voting systems it is necessary to protect. Indefinitely the visitor's privacy even after the election is over, since in case any intruder obtains a digital copy of records that allow the voter to relate to their vote would have all the time to try to decipher it.

People want to keep their privacy secured indefinitely and there are cases in which it would be of the utmost seriousness to know who a person voted. For example, knowing the trajectory as a voter of a current candidate could influence the electorate.

During the electoral process the data security lasts: the protection of the circulating information should only support the period that corresponds to the voting process[8].

Organizations that wish to establish the self in a globalized and competitive market such as today's should have an infrastructure that allows them to interact with their environment in an appropriate way, facilitating the promotion, dissemination and/or provision of their products or services through such a technology platform. However, the destination of such disclosure, processing and/or information processes are not always directed towards the external scope of the organization, but can and should be implemented into the internal processes of the organization, in order to streamline the channels of process control, information (formal and informal), disclosure of policies and feedback of those nerve elements in the organizational work. Despite all the efforts, Charlemagne, in his studies shows that, despite the new channels and digital media, the basic interaction between representatives and represented is constantly relegated[9].

One of the relevant aspects during and after the voting process, is to save the data, for this Legorreta, points out that the transformations that society has suffered over time have led to the raising of new questions about the way in which many of the processes are carried out within a State[10].

The electoral sphere has not been exempted from these transformations. Being the State, who must ensure that the processes are handled efficiently and responsibly.

According to Del Monte, the use of technological resources during an electoral process involves three specific areas:

- Creation of the voter database.
- Total calculation of votes and total aggregation of votes and transmission of results
- Publication of votes and partial results[11].

While the counts have changed, they have influenced using technology has allowed the creation of new means of voting, where technologies are active participants, their foray being for some years in different countries. Manual counting times now involve the use of technologies for electoral management, although this adds benefits in speed, this implies a high cost, both economically as long-term and socially and politically.

In countries like the United States with regular electoral processes, they have already been introduced with equipment that is used to collect votes and then be counted automatically. For this, laws and a jurisdiction have been created that envisages a very rigid state policy.

According to Morales, in that line, people's adaptation to communication and information technologies has been mediated by the level of internet access and the broadband capacity available to them. Thus, the presence of electoral processes in the virtual space depends not only on the characteristics and transformations of web 1.0 to web 2.0, but also the possibility of internet access[12].

There are in turn protocols that have a universal acceptance at the moment to transmit information or data on the Internet, it aims to effectively transmit the information, in addition, by carrying out automated processes, it will allow to expand its performance so that this is implemented within the technical area under way:

- Reduce investment in raw materials, along with human resources.
- Clarity and certainty in the choice of authorities.
- Create favorable conditions for access to suffrage, through conditions that make it easier for the user to comply with their right safely.
- Process information requirements quickly and efficiently.

In addition, Ávila Barrios points out, So far commented that it can be considered an error to address the issue of Electronic Government (GE) only from a technocratic perspective, since it is about having, through the use of technologies, a platform that allow to successfully implement participatory processes, with efficient and transparent procedures that provide effective services to citizens[13].

The information handled in an electoral process would maintain the use of security processes, with data being that are compromised and vulnerable, can be seen in Table 2.

Table 2: Data that Require Security

	Confidentiality	Integrity	Availability
Electoral Roll	NOT	YES	YES
Votes	YES	YES	YES
Candidates	NOT	YES	YES
Income of result	YES	YES	YES

It is a difficult task, but steps are being taken to improve the electoral system and this requires planning, investment and technological updating, as well as Aguilar, notes that the application of new technologies in electoral matters has increased worldwide in the last years[14].

III) Cryptography in Information Security

Information entered an electoral process is the main and most important element in an electoral process, it is sensitive information that must be protected and protected.

Any new approach that comes up is quickly made available to a huge critical mass that evaluates and generates the changes it deems necessary. Simultaneously, the geometric growth of a cryptanalyst's attack capacity requires a cryptographic system to demonstrate security in an undisputed manner, with rigorous formal mathematical techniques to be applied.

Thus, the author García believes that the same is true in the field of computer security in general and in cryptography. One of the reasons for this phenomenon is the large increase in the volume of information available.

- One of the applications with the highest demands in this electronic sense. The results of a vote define important power relations and the management of important economic resources. It is therefore essential to ensure two key points
- The count must transparently reflect the will of the citizens.
- A voter must be assured that their vote will remain anonymous indefinitely[15].

Cryptography in the scrutiny system is used to protect the data transmitted between the voter and the server to ensure that it is not leaked to a third party[16].

The National Electoral Council, as an important point mentions the obtaining of hash codes, this sense, the Hash cryptographic function, is a mathematical algorithm that transforms any arbitrary block or data entry into a series of alphanumeric illegible characters[4].

The Hash algorithm will ensure the integrity of the information due to a system that prevents the manipulation of the results. The identification mechanisms of the users must also be considered, in relation to entering the system with their password, which prevents malicious third parties from wishing to manipulate the system, even preventing these subjects from reversing any procedure already performed by the user. In addition, you can use the digital signature that adds to the user's security system, using encryption with keys that can only be used by authorized persons.

For the cryptographic systems to work, a protocol is required, which allows to follow the respective steps, which avoids the manipulation of some external entity and which allows the proposed objectives to be achieved. Do not underestimate any level of security, so you must be aware of any weaknesses presented and any steps that are not being met accordingly.

IV) hreats in Information Systems

Voting systems even though they have high security standards, however, they can be breached and present certain threats. Among them are:

- Human Threats.
- Natural Threats.
- Environmental Threats.

In the face of the most common threats, several previous practices could be carried out:

- Preparation of Security Protocols.
- Review of Information Security Standards.
- Creation of a Computer Emergency Response Team (CERT).
- Review of Information Security Risk Management Methodologies.
- Review of annual information security reports.
- Mitigate Distributed Denial of Service (DDoS) attacks on Linux platforms.
- Avoid social engineering attacks like spear phishing, which correspond to malicious emails that take users to fake websites with lots of Malware.

According to Camana, the increase in the volume of information that is computerized in digital databases has grown dramatically in recent years[17].

Documentation about the projects to implement automation in the elections, which aims to counteract human error during the entire electoral process, could also be reviewed.

3. Results

Technological suggestions and security measures will enable better counting in a voting process. Abdala says the adoption of electronic voting may have differential effects on election results by affecting voter turnout[18].

In this way, it should be borne in mind that proper handling of information is indispensable, since, when reaching the wrong hands (it can be altered, stolen, deleted, etc.).

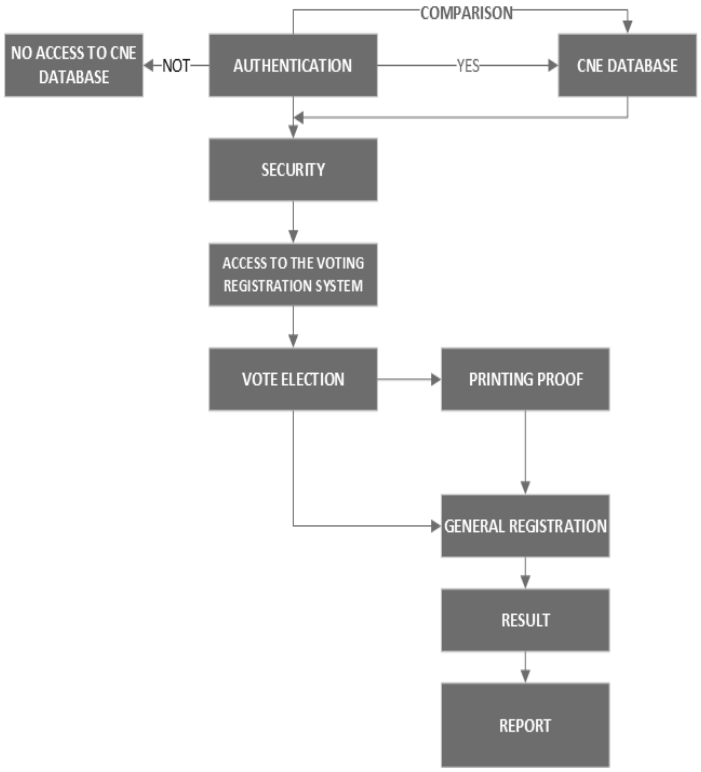


Figure.3 Authentication cryptographic Process

It is also suggested to work with audits to enable it to know whether the mechanism used complies with the security processes and the information entered is being protected and the results are being correct and timely in the voting process, so Rocha, states that, in electronic voting systems, it is extremely important to have a record of events, in order to have the evidence of actions taken especially for cases where manipulations are suspected[19].

- User Authentication.
- Information is encrypted using a cryptographic channel or method.
- If the user complies with security permissions and accesses, the decryption information is verified by security methods.
- If the authentication is successful, the voting registration system is accessed.
- The user casts their vote. Which allows you to print the voting receipt.
- Once the above steps are approved, the general log is generated, its result, and at the end of the report.

It should be added that, the elements that can cause problems to the computer system can be either external (hacker attacks, viruses, espionage) or internal, which can turn out to be very expensive, since, it can be carried out by someone trusted and have direct access to the most important and sensitive information.

Following these considerations, it is possible to avoid computer risks, but also by taking care of possible accidents of nature that may affect processes and through the creation of public policies, establishing rules and procedures that allow safeguard election information.

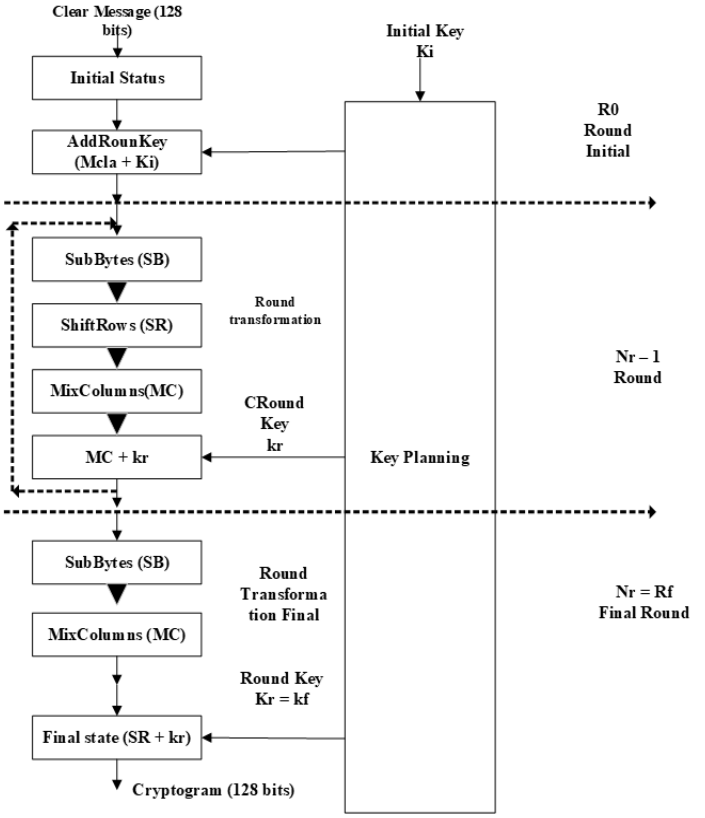


Figure.4 Cryptography AES

It was determined that the combination of the symmetrical system and the asymmetric system can be combined to obtain a better result.

This hybrid security system will prevent external agents from manipulating or affecting the results obtained in the different elections; complementary safety measures have also been analyzed that can help mitigate or avoid possible effects on the results.

The AES encryption algorithms, consists in applying to each state a set of grouped operations that are considered rounds, the algorithm performs eleven rounds, where in each round a different subkey is applied:

- 1 Initial Round.
- 9 Rounds of Transformation.
- 1 Final Round, as seen in the figure 6

4. Discussion

It can be understood that new technologies contribute effectively to progress in each country, in Latino America there are few countries that use electronic voting, Ecuador still does not specify this proposal because of a lack of decision, budget or distrust in political parties. With a good programmed established and following security protocols, its implementation could ensure security, speed and effectiveness, before, during and after the electoral process. Any voting system, whether digital or traditional, will need essential requirements for its implementation. However, there is no voting system that provides this security and integrity of information; more, however, there are aspects that can be taken into consideration to prevent certain attacks.

Hence it is suggested to work with a permanent audit, as noted, Schmidt, the functions of the system should be stored in digital logs that can be extracted later for audit purposes. These blogs can also be used as duplicates of the documents required by current legislation[20].

But it is also suggested to take brief steps to develop an appropriate security system:

- It is necessary to recognize the required infrastructure and the proper security measures both physical and digital, as well as the care of those who have access to such information and the custodians who provide security of the site.
- Study the vulnerabilities that can occur within the system and the enclosure where such data is permanently located, as well as external attacks that attempt or violate security measures.
- Develop a contingency plan in case security measures are breached.
- Have a team of professionals dedicated to maintaining stability and security in networks and identifying potential vulnerabilities, as well as malicious attacks, fixing open door problems.
- Establish security policies both in the use of the system and in the responsibilities in case of eventualities during its application.

Security systems that are applicable in some contexts have been noted that they do not have the same effect on others so models vary in countries that use electronic voting systems in whole or in part, as is the case in Argentina compared to Venezuela, which apply different models. Thus, the suggestions made for Ecuador may be effective if other complementary aspects are met.

Despite the contrary intentions, from some malicious and direct sources to the electoral system, these suggestions will mitigate external interventions that want to change the election results.

5. Conclusions and Future Works

5.1. Future Works

Design of security protocols, new cryptographic methods and effective physical and logical security measures to prevent information leakage and modification of the results obtained in the different votes, thus avoiding problems with the information entered in the electoral process.

Strategies for the protection of precincts before, during and after sectional voting, using advanced technology.

Technology security audit, which allows monitoring the compliance of the phases during the electoral processes.

5.2. Conclusions

In the voting processes for Sectional Governments in Ecuador and requires working with a design of security protocols with strategies that counter social engineering attacks by preventing the leakage of information and modification of the results obtained in the votes.

It is necessary to use new technologies so that the results obtained are reliable, it is pertinent that the cryptography implemented by the CNE with the hash code is in constant integration with new more robust and secure cryptographic methods to maintain an adequate security standard, since it allows a correct encryption of the information, providing greater security, integrity and confidentiality to the corresponding votes, likewise care must be taken of content uploaded in clouds or on state government platforms, because they can be vulnerable to specific attacks by malicious codes or by denial of service.

Finally, to involve international entities which allow audits to be carried out and help improve the current state of information security systems in the country.

Acknowledgment

The authors thank the University Politécnica Salesiana of the Ecuador, to the research group of the headquarters of Guayaquil "Computing, security and Informatics for a globalized world" (CSITGW) created in accordance with resolution 142-06-2017-07-19 and the Secretariat education higher science, technology and innovation (Senescyt).

References

[1] C. D. De Faveri, A. Moreira, J. Araújo, and V. Amaral, “Towards security modeling of E-voting systems,” Proc. - 2016 IEEE 24th Int. Requir. Eng. Conf. Work. REW 2016, pp. 145–154, 2017.

[2] J. Rojas-mora, "Desafíos jurídico-políticos en el entorno digital Sulan Wong Ramírez Julio Rojas-Mora Helder Binimelis Espinoza," vol. 28, no. 2015, pp. 8–12, 2018.
[3] A. del Ecuador, "Constitución del ecuador," 2008.
[4] J. Badí and Q. Basantes, "Análisis de la seguridad integral en los procesos electorales : Caso proceso electoral 2017 de Ecuador," pp. 105–130, 2017.
[5] Telefonica, Ciberseguridad , la protección de la información en un mundo digital. Madrid, España: 2016, 2016.
[6] G. Gallegos-garcía, R. Gómez-cárdenas, and G. Duchén-, "Protocolo de votación electrónica basado en emparejamientos bilineales Electronic voting protocol from bilinear pairings," pp. 234–244, 2010.
[7] A. A. TOSSI, "Revista Política y Estrategia," vol. 125, 2015.
[8] S. Bast, P. García, and G. Montejano, "OTP-Vote : Avances en la Generación de un Modelo de Voto Electrónico," pp. 1069–1073.
[9] M. C. Carlomagno, S. S. Braga, and R. C. Sampaio, Respondem os políticos a questionamentos dos eleitores ? Um experimento controlando os incentivos de mensagem , período e meio, vol. 24. 2018.
[10] P. Carolina, G. Legorreta, P. Carolina, and G. Legorreta, "Voting from Abroad: What Did We Learn From the 2006 Experience?," 2014.
[11] F. B. del Monte, Límites y potencialidades del Voto Electrónico. 2007.
[12] J. A. Morales and J. Alexander, "El impacto de la interacción virtual en los procesos electorales en México," vol. 9, pp. 9–45.
[13] D. A. Barrios, "La implementación de Tecnologías de la Información y Comunicación (TIC) desde finales del siglo pasado produjo , a escala mundial , importantes cambios en todo ámbito , sea privado o público , desde la forma de organización personal e institucional hast," 2014.
[14] D. E. Aguascalientes, A. Lilia, S. Aguilar, C. Gutiérrez, L. Cristina, and P. Howlet, "Electronic voting : reliability and implementation of technology," pp. 77–83, 2017.
[15] P. García, G. Montejano, and S. Bast, "Seguridad Incondicional para el Anonimato en Sistemas de e-Voting," 2015.
[16] L. Rura, B. Issac, and M. K. Haldar, "Secure electronic voting system based on image steganography," 2011 IEEE Conf. Open Syst., pp. 80–85, 2011.
[17] R. Camana, "Potenciales Aplicaciones de la Minería de Datos en Ecuador Minería de datos : Orígenes y Conceptos," vol. 29, no. Julio, pp. 170–183, 2016.
[18] M. Belén, A. Benesch, and A. Pedro, "Evaluación de los efectos de la Boleta Única Electrónica : evidencia experimental de las elecciones en Chaco 2015 * Evaluation of the effects of e-voting : Experimental evidence," 2016.
[19] L. Felipe and F. Martínez, "Audit Mechanism for Detecting Voting Manipulation in Electronic Voting Systems Audit Mechanism for Detecting Voting Manipulation in Electronic Voting Systems Óscar Ruiz Hernández **," no. January 2018, 2019.
[20] J. Hacia, "Hacia el desarrollo de un prototipo de sistema de voto electrónico para Costa Rica Towards the development of a electronic voting system prototype for the Costa Rican context," vol. 29, pp. 146–158, 2016.

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 360-368 (2019)

www.astesj.com

ASTESJ

ISSN: 2415-6698

LED Lighting Implementation as a Strategy for Energy Saving in Energetically Sustainable Campus

Pablo Pérez-Gosende[1,2,*]

[1]Grupo de Investigación Interdisciplinar en Matemática Aplicada (GIIMA), Universidad Politécnica Salesiana, 090109, Ecuador

[2]Centro de Investigación en Gestión e Ingeniería de Producción (CIGIP), Universitat Politècnica de València, Alcoy, 03801, Spain

ARTICLE INFO

Article history:
Received: 04 August, 2019
Accepted: 13 September, 2019
Online: 22 October, 2019

Keywords:
AHP
Dialux
Energy-saving strategy
LED lighting

ABSTRACT

Since lighting constitutes an important part of the total electricity consumption of a country, the design and implementation of more efficient interior lighting systems are one of the feasible strategies to reduce energy consumption. With this purpose, in recent years several studies have considered the use of LED lighting systems. However, its implementation at a large scale is still unaffordable for some organizations like academic institutions. In this context, this research presents a strategy to replace the current fluorescent-based lighting system to LED technology in a building of a higher education institution in Ecuador, and thus contribute to reducing energy consumption in campus facilities. The appropriate LED lamp alternative was selected from those available in the Ecuadorian market through AHP multicriteria decision-making method. The improvement proposal included an annual schedule for the progressive replacement of fluorescent lamps, based on the economic savings that would be generated every year after its implementation. The proposal was technically validated by simulation in DIAlux 4.12 and its economic feasibility was demonstrated by estimating the net present value and the equivalent uniform annual cost. With this strategy, a total of 2595 fluorescent tubes installed in the 215 building light points, can be replaced with LED technology in three years, with an initial investment of 4949.10 USD and generate an energy saving of 47.36%.

1. Introduction

Given the high level of pollution generated in the environment by the production of electricity based on fossil fuels, it is evident that there is a need to draw up strategies to reduce energy consumption. For this purpose, one of the feasible strategies consists in the design and implementation of more efficient interior lighting systems [1], since lighting constitutes an important part of a country's total electricity consumption [2]. It has been reported that in industrialized countries, lighting represents 20% of the total energy consumption [3].

The optimization of indoor lighting systems is a matter of great interest to the scientific community [4]. Approximately 21% of all the electricity produced in the world is consumed in lighting [5]. Some more specific studies show that the electric energy consumed by lighting systems represents 50% of the total energy consumption in office buildings, between 20% and 30% in hospitals, 15% in manufacturing factories and from 10 to 15 % in schools [4,6].

In the USA approximately 31% of all electrical consumption in academic buildings is due to lighting systems [7]. In Europe, this value is approximately 50% [8]. The reduction of energy consumption through the design and implementation of more efficient lighting systems is an important solution to the problem [5], but not the only one. Other possible ways to reduce the energy consumption related to lighting in buildings include the creation of awareness in the users, the adjustment or organization of the furniture and other elements within the space in a way that allows to take advantage of the natural light that enters through the windows, or the installation of motion sensors and control systems that modulate the intensity of the light according to the needs or turn it off when the space is not being occupied.

In recent years, several studies have considered the implementation of LED lighting (Light-Emitting Diodes) to reduce the electricity consumption of the facilities and thus contribute to the energy efficiency in buildings [1,4,5,9,10]. LED lighting has gained special relevance in recent years over traditional lighting technologies due to its high efficiency, long

*P. Pérez-Gosende, Chambers 227, Guayaquil, Ecuador, pperezg@ups.edu.ec

lifespan and low maintenance cost. However, its impact on energy savings in industrialized countries has been questioned [3].

LEDs can produce high luminous fluxes with low heat emission and maintain their light output efficiency for years. Incandescent (IL) and fluorescent (FL) bulbs contain filaments that must be replaced periodically and consume a lot of electrical energy while generating heat [11]. The duration of the fluorescent bulbs is of the order of 20,000 hours, while the incandescent bulbs last approximately 1000 hours. LEDs, on the other hand, do not have filaments, and therefore, they do not burn like incandescent or fluorescent bulbs. Normally, commercial LED lamps have a lifespan of 30,000 to 50,000 hours.

The design of an interior lighting system should consider, in addition to energy efficiency, the visual performance and comfort of users [5]. Particularly in educational institutions, the implemented lighting systems should provide a pleasant and stimulating environment that allows students, teachers and administrative staff in general, to perform their activities without excessive visual effort, avoiding fatigue, headaches and the prevalence of vision disorders provoked by inadequate lighting [12]. For this reason, the lighting system in educational institutions should be monitored with a certain frequency to guarantee the correct development of the teaching-learning processes.

In Ecuador, the 2393 Executive Decree that regulates the levels of lighting is not very demanding, or at least not very precise in terms of the levels of lighting suitable for each activity [13]. On the other hand, public or private institutions are not required to carry out light audits to detect deficiencies in lighting systems. Therefore, it is extremely important to give concrete answers to the reduction of energy consumption in Higher Education Institutions (HEIs) by searching for more efficient lighting alternatives without neglecting compliance with current regulations regarding health and safety.

In this context, this study proposes an energy-saving strategy based on the progressive implementation of LED lighting in an academic building of an HEI located in the city of Guayaquil.

This research is part of a wider scope project named "Sustainable Energy Campus Model" developed by the Interdisciplinary Research Group in Applied Mathematics (GIIMA) of the Universidad Politécnica Salesiana in conjunction with the Interdisciplinary Modeling Group (InterTech) of the Universitat Politècnica de València. The results of the project will allow, in the medium term, to optimize the energy consumption for lighting, air conditioning and the use of computers in the institution. The future savings generated by the improvements proposals within this project could further strengthen the university infrastructure, especially in the setting and equipment of new laboratories and the increase and updating of the literary heritage available in the university library.

2. Methods

The research is quantitative with a descriptive approach. The HEI under study is located in the city of Guayaquil, Ecuador, and is composed of six buildings. These, in turn, are composed mostly of classrooms, laboratories, teacher workrooms, administrative offices, auditoriums, hallways, bathrooms, and storage areas.

Illuminance measurements were taken in the facilities of one of these buildings between October 2017 and February 2018 by four AMPROBE lux meters model LM100, with an accuracy of ± 5% + 5 digits. The dimensions of the facilities, as well as the distance from the work surface to the light points, were obtained through a Capital laser distance meter, model CP-3007, with an accuracy of ± 0.5% and resolution of 0.01m.

The replacement strategy of the current lighting system based on FL to LED technology, included an annual schedule for the progressive replacement of fluorescent lamps, based on the economic savings that would be generated every year after its implementation. For this, the LED lamp alternative suitable for the building's facilities was selected from those available in the Ecuadorian market using the AHP multicriteria method [14].

For the improvement proposal technical validation, it was used DIALux 4.12 software. Facilities were modeled and the lighting levels that would be obtained once the lamps were installed were simulated. In this regard, it was possible to verify in each case the fulfillment of the minimum illuminance requirements established in the national and international regulations used as a reference in this work [12,13].

For the economic validation of the improvement proposal, using (1), the net present value (NPV) of the costs was calculated and compared for the two project alternatives: (A) The progressive replacement to LED lighting or (B) to keep the current FL-based lighting system.

$$NPV = \sum_{j=1}^{n} PV_j \quad (1)$$

By progressively replacing the current FL-based lighting system with LED technology, energy consumption will decrease and this will generate economic savings. This saving has been taken into account in this investigation for the determination of the annual present value (PV_j), considering as such the updated value of the difference between the total costs (TC_j) and the future annual benefits (ES_j) as presented in (2).

$$PV_j = (TC_j - ES_j) \cdot [1/(1+i)^j] \quad (2)$$

Likewise, the economic analysis included the equivalent annual uniform cost estimation (EAUC) for each alternative throughout its expected time horizon. EAUC consists of converting all expenditures into a uniform series of payments [15] and is frequently used in the evaluation of investment projects where revenues are not relevant, but costs. It is formulated as in (3) and its value depends on the capital cost or discount rate (i) and the NPV of the amounts of cash outflows from year j=0 to year j=n. When comparing several alternatives of mutually exclusive projects, the one with minor EAUC should be selected [15,16].

$$EAUC = NPV\ [i(1+i)^n/(1+i)^n - 1] \quad (3)$$

The relevant benefits and costs items for NPV calculation (expressed in $/year) are obtained from the following equations. There, AC represents the annual cost of lamps acquisition, EC the annual electricity consumption cost, DC the cost of the final disposal of the fluorescent lamps that are replaced and ES corresponds to the annual economic benefit that represents the energy savings by the use of LED lamps instead of FL. In addition, L represents the quantity of lamps expressed in units (u), Pr the wholesale price of the lamp ($/u), P the power of the lamp (W), H

the estimated daily use time of the lamp, D the number of days per year in which the institution operates and T is the electric rate ($/kWh). Also, W_{FL} refers to the average weight of a 60cm long fluorescent lamp (Kg/u) and PrD means the price of the specialized recycling service for the final disposal of FL ($/Kg).

$$TC_j = AC_j + EC_j + DC_j \quad (4)$$

$$AC_j = L \cdot Pr \quad (5)$$

$$EC_j = 0.001 \cdot L \cdot P \cdot H \cdot D \cdot T \quad (6)$$

$$DC_j = W_{FL} \cdot L_{FL} \cdot Pr_D \quad (7)$$

$$ES_j = 0.001 \cdot L_{LED} \cdot H \cdot D \cdot T \cdot (P_{FL} - P_{LED}) \quad (8)$$

3. Results

In this section it is reported the results of the light audit, the selection of the most suitable LED lamp alternative for educational facilities according to certain relevant factors, the proposed replacement schedule and the technical and economic validation of the improvement proposal.

3.1. Indoor lighting audit

The building under study is one of the six that comprise the campus. There were identified 50 classrooms, 4 hallways, 18 restrooms, 8 computer laboratories, 4 teacher workrooms, 5 administrative offices, a medical department, a supply warehouse, a restaurant/coffee shop and a copy center, all of them illuminated by 60 cm long linear fluorescent lamps with powers of 17, 18 and 32W. 215 light points and 2595 fluorescent tubes were counted throughout the building. The chapel, the theater, and the auditoriums do operate with compact fluorescent lamps (CFL) downlight style.

In the building, several careers carry out their daily academic activities in three sessions: morning (07h00-13h00), evening (14h00-18h00) and night (18h00-22h15). The existence of a night session facilitates the students to work and attend University at the same time and is one of the differentiating elements of this HEI in the city. After the last shift of classes, it takes an hour for the cleaning team to organize all the classrooms for the next day.

According to the above, in this study, it has been considered that the lighting system remains on 16 hours a day, 288 days a year. The illuminance measurements were made at night, between 7:00 p.m. and 10:00 p.m. In each location, illuminance measurements were taken at the height of the work surface (0.77m in classrooms and 0.8m in laboratories) at five different points, and the average value was taken as a reference.

The legislation that regulates indoor and outdoor lighting levels in Ecuador is the 2393 Executive Decree: Regulation of safety and health of workers and improvement of the work environment [13], however, this norm does not establish specific illuminance values for academic activities given its generality. From its content, it is assumed that the reading and writing activities that take place in a classroom, as well as those that take place in a computer laboratory, can be done with a minimum lighting level of 300 lx since in these cases the average distinction of details is essential. In these facilities would not be justified an illuminance of 500 lx, which is the value established in the standard for works in which a fine distinction of details under conditions of contrast is indispensable.

Given the lack of specificity in the Ecuadorian regulations, this research also took as a reference to the European standard on indoor lighting EN 12464.1, which includes minimum illuminance values for educational establishments [12]. Table 1 shows a comparison of the levels of illumination allowed for the facilities at the academic building under study in this investigation according to both regulations. In case of disparity, the highest minimum permissible illuminance value was taken as reference.

According to the above, it was identified that 98% of the building facilities do not comply with the lighting requirements established in the national regulation and the European standard for indoor lighting. Note also that only the chapel and the restaurant/coffee shop meet the reference standard used in this investigation.

The noncompliance with the lighting levels required for the facilities was corroborated through its simulation in DIAlux 4.12. Taking as an example a standard classroom model of 6.80m x 6.70m x 3.25m, considering a maintenance factor of 0.8, a height of the work plane of 0.77m, six light points and three FL lamps of 17W in every light point, the resulting illuminance isoline diagram (Figure 1) shows illuminance values below the 300 lx requirement established in regulations for this type of facility.

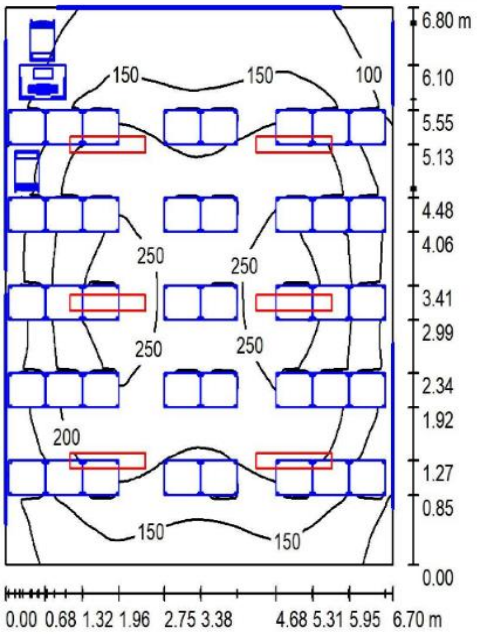


Figure 1: Illuminance isoline diagram for a standard model classroom

3.2. LED lamps selection

To verify more economical lighting alternatives and contribute to the campus energy saving, it was decided to implement LED lighting, as this technology is more efficient, generates less consumption given its low power, in addition to its longer lifespan.

Unified Glare Index, UGR), efficiency (lm/W) and light type. Then, as all the identified alternatives provided light in a neutral white tone and had a UGR<19, the criteria type light and glare were discarded from the final decision process.

Table 1: Compliance with illuminance reference standards for academic buildings

Facilities	Standard		Reference standard	Total of facilities	Facilities that do not meet the standard	Measured illuminance		Illuminance needed to meet the standard	
	Decree No. 2393	EN 12464.1				μ	σ	μ	σ
Classrooms	300 lx	300 lx	300 lx	50	50	130	46.30	170	46.30
Hallways	20 lx	100 lx	100 lx	4	4	54	12.53	47	12.53
Restrooms	50 lx	200 lx	200 lx	18	18	35	8.03	165	8.03
Computer laboratories	300 lx	300 lx	300 lx	8	8	139	36.60	161	36.60
Teacher workrooms	Not specified	300 lx	300 lx	4	4	112	38.24	189	38.24
Administrative offices	300 lx	300 lx	300 lx	5	5	113	33.68	187	33.68
Medical department	Not specified	500 lx	500 lx	1	1	225	0	275	0
Warehouse	200 lx	100 lx	200 lx	1	1	140	0	60	0
Restaurant and coffee shop	Not specified	[a] Not specified	100 lx	1	0	260	0	0	0
Chapel	Not specified	Not specified	100 lx	1	0	178	0	0	0
Copy center	300 lx	300 lx	300 lx	1	1	145	0	355	0
Theater	Not specified	Not specified	100 lx	1	1	84	0	16	0
Auditoriums	Not specified	500 lx	500 lx	2	2	169	43	331	43

[a] For this case, the European standard states as appropriate any value of illuminance that allows creating an appropriate atmosphere. Source: Own elaboration based on the requirements in the national and European standards referred to in [12, 13].

The decision criteria that were taken into account for LED lamp selection were: power (W), lifespan (h), wholesale price (USD), color rendering index (CRI), glare (measured by the

The hierarchical decision model followed in this research for the selection of the LED lamp by the Saaty Analytical Hierarchic Process (AHP) is shown in Figure 2. In the scheme, the zero level represents the decision objective, the level one the decision criteria considered, and level two presents a selection of the few alternatives of linear 60cm long LED lamps available in the Ecuadorian market that could be installed in the bases of existing fluorescent luminaires. This last consideration was necessary to reduce the cost of the replacement process, since LED luminaires, despite their benefits, currently continue to be more expensive than FL.

The comparison of the importance or preference between each pair of criteria using the Saaty scale [14], led to a reciprocal matrix (R) so that rij represents the relative priority between criterion Ci and criterion Cj with respect to the objective or goal of the problem.

$$R = \begin{bmatrix} 1 & 5 & 3 & 5 & 1 \\ 1/5 & 1 & 1/3 & 5 & 1/5 \\ 1/3 & 3 & 1 & 5 & 1 \\ 1/5 & 1/5 & 1/5 & 1 & 1/5 \\ 1 & 5 & 1 & 5 & 1 \end{bmatrix}$$

The results of the normalization of the paired comparisons are presented in the R_{NORM} matrix and the resulting priority vector in the W matrix.

$$R_{NORM} = \begin{bmatrix} 0.3659 & 0.3521 & 0.5422 & 0.2381 & 0.2941 \\ 0.0732 & 0.0704 & 0.0602 & 0.2381 & 0.0588 \\ 0.1220 & 0.2113 & 0.1807 & 0.2381 & 0.2941 \\ 0.0732 & 0.0141 & 0.0361 & 0.0476 & 0.0588 \\ 0.3659 & 0.3521 & 0.1807 & 0.2381 & 0.2941 \end{bmatrix}$$

$$W = \begin{bmatrix} 0.3585 \\ 0.1002 \\ 0.2092 \\ 0.0460 \\ 0.2862 \end{bmatrix}$$

The process of paired comparisons of the criteria yielded a consistency index of 9.08%, so it is considered that the matrix R has an admissible consistency and the vector of priorities obtained (W) is accepted as valid.

As could be interpreted in Figure 2, all the decision criteria are quantitative variables, so it would not be relevant to make comparisons between alternatives concerning each variable using the subjective scale of Saaty. Consequently, with the values identified for each decision variable in each alternative, the M

matrix was generated. Here, every mij represents the value that the lamp alternative i takes in variable j.

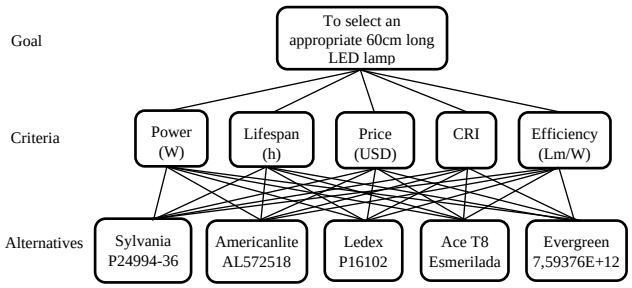


Figure 2: Hierarchical decision model for LED lamp selection

$$M = \begin{bmatrix} 9 & 25000 & 3.70 & 70 & 89 \\ 9 & 40000 & 3.90 & 80 & 130 \\ 9.5 & 30000 & 3.50 & 75 & 77 \\ 9 & 25000 & 3.60 & 75 & 100 \\ 10 & 30000 & 3.60 & 75 & 90 \end{bmatrix}$$

Note that the above mentioned power and wholesale price criteria are to be minimized, while lifespan, color performance, and efficiency are to be maximized. Thus, it was necessary to homogenize the M matrix to the MNORM matrix. The latter is composed of all the local priority vectors of the alternatives for each criterion.

$$M_{NORM} = \begin{bmatrix} 0.1 & 0.625 & 0.0513 & 0.8750 & 0.6846 \\ 0.1 & 1 & 0 & 1 & 1 \\ 0.05 & 0.75 & 0.1026 & 0.9375 & 0.5923 \\ 0.1 & 0.625 & 0.0769 & 0.9375 & 0.7692 \\ 0 & 0.75 & 0.0769 & 0.9375 & 0.6923 \end{bmatrix}$$

Then, the global priority vector of the alternatives was obtained from the multiplication of the alternative priority vectors concerning the decision criteria (MNORM matrix) with the priority vector of the criteria with respect to the objective (W matrix).

$$M_{NORM} \cdot W = \begin{bmatrix} 0.3453 \\ 0.4681 \\ 0.3271 \\ 0.3778 \\ 0.3324 \end{bmatrix}$$

From here it is concluded that the most suitable LED lamp alternative according to the previously defined decision criteria is the Americanlite AL572518. This LED tube, measuring 60cm in length, emits neutral white light, has a power of 9W, 40000h of lifespan, wholesale price of $ 3.90, IRC equal to 80, UGR<19 and average efficiency of 130 lm/W.

3.3. Technical validation of the selected lamp

Taking into account the selected lamp model, the facilities were modeled and the lighting system was simulated using the DIAlux 4.12 software.

Taking the simulation of a standard classroom as an example and considering three LED tubes in each of the six pre-existing light points, the resulting illuminance isoline diagram shows values equal to or greater than those established in the regulations for this type of facility (300 lx) in most of the useful area of the classroom (Figure 3). More specifically, the Dialux report showed an average illuminance value on the work plane of 326 lx, 204 lx on the floor, 70 lx on the ceiling and 52 lx on the walls.

On the other hand, the resulting color performance diagram is shown in Figure 4. The red color on the work surfaces ensures that if the selected LED lamp is implemented, without having to increase the number of light points, it is possible to guarantee the minimum illuminance requirements established in the regulations. This validates from the technical point of view the suitability of the selected lamp for the building's facilities, which are mostly classrooms with similar characteristics.

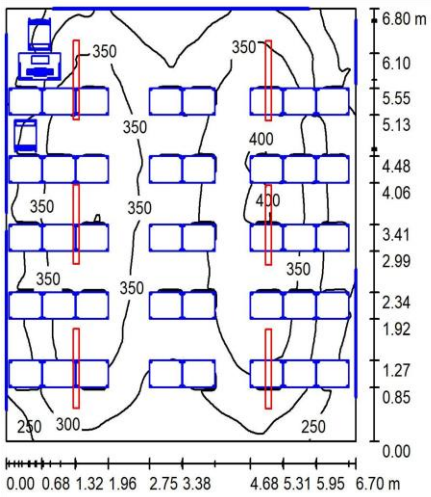


Figure 3: Illuminance isoline diagram for a standard model classroom

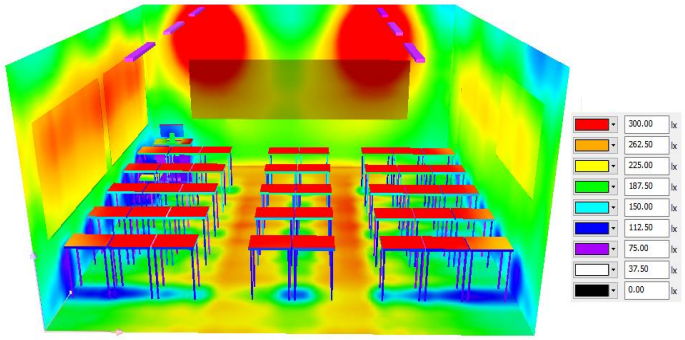


Figure 4: False color rendering

3.4. Replacement strategy

Given the high cost that comprises to shift all the fluorescent lamps at once, it is recommended the replacement to be gradual, prioritizing those facilities with greater gaps to meet the lighting standards. For the building under interest in this research, it is proposed to follow the lamp replacement schedule represented in Table 2 with their respective prioritization criteria. As may be seen, the classrooms and computer laboratories were prioritized, since within these facilities are commonly developed the academic activities that are the central axis of the University formative mission.

Table 2: Schedule for replacement of FL by LED lamps

Priority level	Facility	Quantity	[a] Priority criteria	Required lamps	Schedule
1	Classrooms	15	(RS – IM) ≥ 200 lx	405	First year
2	Computer Laboratories	7	(RS – IM) ≥ 100 lx	189	
3	Classrooms	25	150 ≤ (RS–IM) < 200 lx	297	Second year
4	Computer laboratories	1	(RS – IM) < 100 lx	27	
5	Classrooms	10	(RS – IM) < 150 lx	270	
6	Medical department	1	-	12	
7	Teacher workrooms	4	-	144	
8	Administrative offices	5	-	135	
9	Auditoriums	2	-	78	
10	Hallways	4	-	324	
11	Warehouse	1	-	3	
12	Copy center	1	-	18	
13	Theater	1	-	49	
14	Restaurant and coffee shop	1	-	12	
15	Chapel	1	-	38	Third year
16	Restrooms	18	-	216	

[a] RS represents the value of the reference illuminance standard according to regulations. IM refers to the actual average illuminance measured in the facilities. In areas where no prioritization criteria were applied, all lamps will be replaced at the same time.

It is worth noting that as part of the replacement strategy and after the first year of the initial investment, it has been considered that the new lamps to be installed are to be obtained with the amount equivalent to the economic value saved by the institution the previous year due to the use of LED technology instead of fluorescent technology.

3.5. Economic validation of the improvement proposal

For the economic validation of the improvement proposal, two alternative strategies were considered: (A) To implement the progressive replacement to LED lighting proposed in this study or (B) To continue with the current FL-based lighting system. The energy consumption and annual economic expenditure estimated in each case are referred to in Table 3.

From the above, it is assumed that the current FL-based lighting system electricity costs represent on average 10.93% of the total electrical energy consumed in the building (95% CI: 9.89-11.98), taking into account that during the last 18 months of operation, the total average cost of the building was $ 13,657.94 (95% CI: 12434.50-14881.38).

The investment cycle for strategy (A) and (B), has an estimated time horizon of eight and two years, respectively, considering the average lifespan of the corresponding lamp, the 16 hours of daily use and the 288 days of annual operation of the institution. Then, since both alternatives are mutually exclusive and have different lifetimes, the comparison of their net present values must be made on the least common multiple of years, in this case, 8 years. According to this, Table 4 shows all possible expenditures and benefits for each alternative. In it, for the calculation of the present values of the annual differences between costs and benefits, the active interest rate of 7.26% established by the Central Bank of Ecuador for March 2018 [17] was considered as the minimum attractive rate of return (MARR).

It is important to clarify that in the estimation of the total costs there have not been considered salary expenses of the personnel that will perform the replacement of the fluorescent lamps for those with LED technology, given that the University has maintenance personnel hired and their salaries are debited from the annual operations budget.

With the data in Table 4, a NPV of $56115.02 was calculated for alternative 1, while for alternative 2 the resulting NPV was $133188.18. Then, with the updated values of the differences between costs and benefits, the EUAC was estimated for each alternative considering their respective life cycles. In this regard, the strategy of the progressive switch to LED lighting yielded a EUAC equal to $8868.28, while the alternative of doing nothing and maintain the current FL lighting system was $20669.26. This result shows that the alternative based on LED technology has lower NPV and lower EUAC, which means it is the most feasible alternative from the economic point of view.

Finally, the graph in Figure 5 shows the trend of total costs for both alternatives over time, in addition to the behavior of economic savings and the acquisition cost of LED lamps.

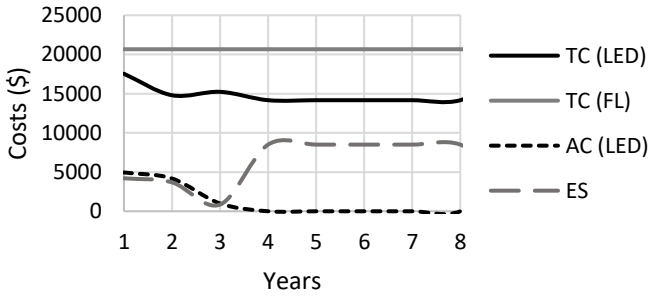


Figure 5: Some relevant cost functions

As can be seen in Figure 5, the strategy of progressive replacement to LED lighting (Strategy a) is less costly during the replacement period (three years) and during the rest of the investment's life cycle. In the same way, it is observed that once the total conversion of the system has been made, the economic saving for the use of LED lighting (with respect to the current system based on FL) would cover the acquisition costs of the new luminaires that are needed at the end of the investment life, guaranteeing the sustainability in time of the proposed system.

Table 3: Energy consumption and annual electricity expenditure for lighting systems based on FL and LED

Facilities	Quantity	Total of lamps	FL Lighting system			LED Lighting system		
			Power consumed (kW)	Power consumption (kWh/year)	Energy cost ($/year)	Power consumed (kW)	Power consumption (kWh/year)	Energy cost ($/year)
Classrooms	50	1350	22.95	105753.60	9517.82	12.15	55987.20	5038.85
Hallways	4	324	5.508	25380.86	2284.28	2.916	13436.93	1209.32
Restrooms	18	216	3.672	16920.58	1522.85	1.944	8957.95	806.22
Computer laboratories	8	216	3.672	16920.58	1522.85	1.944	8957.95	806.22
Teacher workrooms	4	144	2.448	11280.38	1015.23	1.296	5971.97	537.48
Administrative offices	5	135	2.295	10575.36	951.78	1.215	5598.72	503.88
Medical department	1	12	0.216	995.33	89.58	0.108	497.66	44.79
Warehouse	1	3	0.054	248.83	22.39	0.027	124.42	11.20
Restaurant and coffee shop	1	12	0.384	1769.47	159.25	0.108	497.66	44.79
Chapel	1	38	0.684	3151.87	283.67	0.342	1575.94	141.83
Copy center	1	18	0.306	1410.05	126.90	0.162	746.50	67.18
[a]Theater	1	49	0.882	31.75	2.86	0.441	15.88	1.43
[a]Auditoriums	2	78	1.404	808.70	72.78	0.702	404.35	36.39
Total	97	2595	22.95	195247.37	17572.26	23.36	102773.12	9249.58

[a] In the case of the auditoriums, an average use of 2h per day was considered, and for the theater 8h per month, wich is equivalent to 0.3h per day.

4. Discussion

This research shows that the progressive implementation of LED lighting in an academic building is a strategy that can contribute substantially to the energy and economic savings of a Higher Education Institution.

A lighting audit found that 98% of the building's facilities do not meet the illuminance requirements included in national and international regulations [12,13]. Likewise, it was detected that the energy consumption of the current FL-based lighting system represents 10.93% of the total electric energy consumed in the building, a value considered high, despite being located within the range of 10 to 15% identified by other similar studies in educational institutions [4,6].

These reasons justify the need to implement a strategy to optimize energy saving through lighting system in the building, which at the same time, favors the performance and visual comfort of users through compliance with current regulations on occupational health and safety.

In response to this need, this research presents a strategy for the progressive replacement of the current FL-based lighting system to LED lighting, with a duration of three years, which guarantees compliance with the minimum levels of lighting required by national and international regulations for each facility and allows to obtain an energy saving of 47.36%. This value would represent an economic saving of $8322.68 per year for the institution and would allow the total coverage of the new luminaires that are needed at the end of the investment lifecycle.

Since the instantaneous conversion to LED lighting can be unfeasible from the economic point of view due to the high amount of the initial investment, the strategy suggested in this research prioritizes the replacement in those facilities with a greater need for illuminance to meet the requirements on national and European regulations. In this concern, the planned replacement of lamps can be covered with the amount equivalent to the economic value saved by the institution the previous year due to the use of LED technology instead of FL technology. Another element of importance to highlight in order to reduce installation costs was the consideration of the same number of pre-existing light points.

According to the above, the amount of the initial investment of $4949.10 will allow replacing 48.90% of the lighting system. Then, with the energy savings of the first year, 1072 LED lamps could be acquired to increase coverage to 90.21%. Finally, with the energy savings of the second year, the remaining 254 lamps can be acquired to cover 100% of the building's needs according to the planned schedule.

The technical validity of the improvement proposal was obtained by modeling and simulating all the building's facilities in the free software DIALux, which is one of the most commonly used tools in lighting systems simulation [1,4,5,9,10,18–21]. In this regard, this research demonstrates that the results of the simulation in DIALux are very similar to those on-site

Table 4. Costs and benefits of the alternatives: (a) Progressive change to LED lighting and (b) Current lighting system based on FL

Alternatives	[a] Cost ítems	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8
(A) Progressive swift to LED lighting	a	4949.10	4180.80	990.60	0	0	0	0	0
	b	4736.52	8482.99	14167.56	14167.56	14167.56	14167.56	14167.56	14167.56
	c	7465.58	1806.52	0	0	0	0	0	0
	d	390.85	330.18	78.23	0	0	0	0	0
	e	17542.05	14800.49	15236.39	14167.56	14167.56	14167.56	14167.56	14167.56
	f	4210.24	3690.18	858.47	8504.14	8504.14	8504.14	8504.14	8504.14
	g	12429.44	9657.19	11651.52	4278.85	3989.23	3719.22	3467.48	3232.78
(B) Maintain the current FL-based lighting system	a	2334.60	2334.60	2334.60	2334.60	2334.60	2334.60	2334.60	2334.60
	c	17935.19	17935.19	17935.19	17935.19	17935.19	17935.19	17935.19	17935.19
	d	399.48	399.48	399.48	399.48	399.48	399.48	399.48	399.48
	e	20669.26	20669.26	20669.26	20669.26	20669.26	20669.26	20669.26	20669.26
	g	19270.24	17965.92	16749.88	15616.15	14559.15	13573.70	12654.95	11798.39

[a] The cost items considered are: a) Acquisition, b) Annual energy consumption in LED lighting, c) Annual energy consumption in FL lighting d) Cost of the final disposal of FL, e) Total cost, f) Economic saving by the use of LED versus FL and g) Present value of the difference between costs and benefits.

measurements, confirming what was reported in previous studies [10,21].

Several authors have proposed the adoption of LED lighting as an alternative to the reduction of energy consumption in educational facilities [1,4,5,18,21]. However, these studies start from the use of a specific luminaire model, devaluing the analysis of other alternatives through a multicriteria decision method. In this research, the application of the AHP method for this purpose demonstrated feasibility and relevance. In the same way, the lamps selection criteria here used and their respective levels of relative importance may constitute a starting point for future research to strengthen lamp selection decisions.

Another important element to highlight is that the studies that have validated the implementation of indoor LED lighting systems as the best strategy against traditional lighting technologies [1,4,5,9,18], do not consider the costs of the final disposal of a massive FL replacement. Fluorescent and compact fluorescent lamps contain Mercury, a potentially dangerous element for the environment, so the final disposal of these lamps must be handled carefully through special recycling techniques [2,3]. For this reason, in this research when estimating the costs for each alternative, in addition to the acquisition costs and the annual energy consumption, the cost of the final disposal of the replaced FL was included.

In the economic analysis of alternative lighting systems for interiors, it is common to calculate the Simple Payback Period (SPBP) or the whole lifecycle cost (WLC) without considering the value of money in the time [1,4,10,19], although some authors consider that these techniques should not be used as the primary measure of value to select a project alternative [15,16]. The consideration of the net present value (NPV) or the equivalent annual uniform cost (EAUC) as economic evaluation tools is not common in the literature.

In this context, the strategy of progressive change to LED lighting proposed in this research yielded a VPN of $ 56115.02. This means that if the institution had to disburse today the amount required to cover the lighting costs of the next eight years, it would be more economical to have a lighting system based on LED technology since maintaining the current FL-based system would be 42% more expensive. Likewise, for the first strategy, the estimated EAUC was $ 8868.28, less than the value of $ 20669.26 which would represent maintaining the current system based on FL technology. This allows validating that the first alternative is more feasible from the economic point of view, and at the same time confirms the relevance of the VPN and the EAUC as economic evaluation tools for mutually exclusive investment project alternatives with different life cycles where generally only the costs are relevant.

Finally, the authors would like to point out that the reduction of energy consumption in a university campus, not only depends on the implementation of strategies such as the adoption of a more efficient and safe lighting system like the one suggested in this work. Energy-saving must be a shared social responsibility of all those who make the University a common resource for the entire Society. In this regard, HEIs should take advantage of their transformative capacity to raise awareness among their students, teachers and administrative staff about the need to contribute to energy savings and thereby maximize the contribution that is made from the academy towards environmental sustainability.

5. Conclusions

This work showed that despite the high cost of LED lighting, it is feasible from a technical and economic point of view to converting an FL lighting system set in educational institutions through the implementation of a progressive replacement strategy. The suggested strategy guarantees the total replacement of the current lighting system of a university building consisting of 215 light points and 2595 FL, in three years, with an initial investment of 4949.10 USD, generating an energy saving of 47.36%.

Through simulation in DIALux 4.12, it was demonstrated that the proposed LED lighting system will cover the minimum indoor lighting levels established in national and international regulations for educational facilities. The estimation of the investment net present value, as well as its equivalent annual uniform cost, proved the economic feasibility of the improvement proposal.

The results of this work could constitute a reference for the administrative management of public universities in developing countries, which increasingly receive fewer resources from governments. The economic savings for the implementation of

more efficient lighting systems could contribute to the strengthening of university infrastructure and the offering of better educational services. At the same time, the energy savings generated would have a positive impact on the reduction of the emission of greenhouse gases into the atmosphere. In this regard, future research should pay more attention to the combination of more efficient lighting systems with cleaner energy sources, to maximize the contribution that is made from the academy to the sustainability of the environment.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

Authors would like to thank Universidad Politécnica Salesiana for funding this research under the resolution No. 03-006-2017-05-30 of Guayaquil Campus Research Council. Also, a special acknowledge is dedicated to all the industrial engineering students who helped to collect relevant data for the lighting audit in all the facilities of the academic building.

References

[1] N. Tshikuvhe, M. Nthontho, S. P. Chowdhury, S. Chowdhury, "Cost effective carbon mitigation through energy efficiency: A case for a university student residence lighting in Cape Town," in 2013 IEEE African Conference (Africon), Pointe-Aux-Piments, Mauritius, 2013. https://doi.org/10.1109/AFRCON.2013.6757686

[2] M. M. Aman, G. B. Jasmon, H. Mokhlis, A. H. A. Bakar, "Analysis of the performance of domestic lighting lamps," Energ. Policy, 52, 482–500, 2013. https://doi.org/ 10.1016/j.enpol.2012.09.068

[3] B. Gayral, "LEDs for lighting: Basic physics and prospects for energy savings," Comptes Rendus Phys., vol. 18, no. 7–8, pp. 453–461, Sep. 2017. https://doi.org/10.1016/j.crhy.2017.09.001

[4] F. Salata, I. Golasi, M. di Salvatore, and A. de Lieto Vollaro, "Energy and reliability optimization of a system that combines daylighting and artificial sources. A case study carried out in academic buildings," Appl. Energ, 169, 250–266, 2016. https://doi.org/10.1016/j.apenergy.2016.02.022

[5] P. Sathya, R. Natarajan, "Energy estimation and photometric measurements of LED lighting in laboratory," in 2014 International Conference on Advances in Electrical Engineering (ICAEE), Tamilnadu, India, 2014. https://doi.org/10.1109/ICAEE.2014.6838470

[6] B. Atanasiu, C. Despret, M. Economidou, J. Maio, I. Nolte, O. Rapf, "Europe's Buildings under the Microscope: A Country-by-Country Review of the Energy Performance of Buildings," Build. Perform. Inst. Eur., 2011.

[7] J. Min, I. L. Azevedo, P. Hakkarainen, "Assessing regional differences in lighting heat replacement effects in residential buildings across the United States," Appl. Energy, 141, 12–18, 2015. https://doi.org/10.1016/J.APENERGY.2014.11.031

[8] Y. Nikolaidis, P. A. Pilavachi, A. Chletsis, "Economic evaluation of energy saving measures in a common type of Greek building," Appl. Energy, vol. 86(12), 2550–2559, 2009. https://doi.org/10.1016/J.APENERGY.2009.04.029

[9] S. Ji, G. Cao, J. Zhang, F. Yu, D. Li, J. Yu, "Lighting design of underground parking with tubular daylighting devices and LEDs," Opt. - Int. J. Light Electron Opt., 127(3), 1213–1216, 2016. https://doi.org/10.1016/j.ijleo.2015.10.189

[10] P. K. Soori S. Alzubaidi, "Study on improving the energy efficiency of office building's lighting system design" in 2011 IEEE GCC Conference and Exhibition, 2011. https://doi.org/10.1109/IEEEGCC.2011.5752604

[11] D. Singh, C. Basu, M. Meinhardt-Wollweber, B. Roth, "LEDs for energy efficient greenhouse lighting," Renew. Sustain. Energy Rev., 49, 139–147, 2015. https://doi.org/10.1016/j.rser.2015.04.117

[12] Comisión de Normalización Europea, Iluminación de los lugares de trabajo en interiores. Unión Europea, 2002.

[13] Instituto Ecuatoriano de Seguridad Social, Reglamento de Seguridad y salud de los trabajadores y mejoramiento del ambiente de trabajo. Ecuador, 1986.

[14] T. L. Saaty, "Decision making with the analytic hierarchy process," Int. J. Serv. Sci., 1(1), 83–98, 2008.

[15] N. Sapag Chain, R. Sapag Chain, Preparación y evaluación de proyectos, McGraw Hill Interamericana, 2008.

[16] L. Blank and A. Tarquin, Ingeniería Económica. McGraw-Hill Interamericana, 2006.

[17] Banco Central del Ecuador, Información económica, 2018. Available: https://www.bce.fin.ec/index.php/informacioneconomica.

[18] L. Alboteanu, G. Manolea, "Photovoltaic lighting for an educational room," in 2013 4th International Symposium on Electrical and Electronics Engineering (ISEEE), Galati, Romania, 2013. https://doi.org/10.1109/IEEEGCC.2011.5752604

[19] S. Alzubaidi, P. K. Soori, "Energy Efficient Lighting System Design for Hospitals Diagnostic and Treatment Room—A Case Study," J. Light Vis. Environ., 36(1), 23–31, 2012. https://doi.org/10.2150/jlve.36.23

[20] L. E. Mavromatidis, X. Marsault, H. Lequay, "Daylight factor estimation at an early design stage to reduce buildings' energy consumption due to artificial lighting: A numerical approach based on Doehlert and Box–Behnken designs," Energy, 65, 488–502, 2014. https://doi.org/10.1016/j.energy.2013.12.028

[21] K. Zalatis, K. Berzina, "Effective and optimal simulation of light design in Riga Technical university's lighting laboratory from the point of view of energy efficiency," in 2016 IEEE 4th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE), Vilnius, Lithuania, 2016. https://doi.org/10.1109/AIEEE.2016.7821822

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 369-374 (2019)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Verification of the Usefulness of Personal Authentication with Aerial Input Numerals Using Leap Motion

Shun Yamamoto*, Momoyo Ito, Shin-ichi Ito, Minoru Fukumi

Tokushima University, Information Science and Intelligent Systems,770-8506, Japan

ARTICLE INFO

Article history:
Received: 05 August, 2019
Accepted: 04 October, 2019
Online: 22 October, 2019

Keywords:
Aerial input numerals
Leap motion
Deep learning
CNN
Dilated convolution
Personal authentication

ABSTRACT

With the progress of IoT, everything is going to be connected to the network. It will bring us a lot of benefits however some security risks will be occurred by connecting network. To avoid such problems, it is indispensable to strengthen security more than now. We focus on personal authentication as one of the security.

As a security enhancement method, we proposed a method to carry out numeral identification and personal authentication using numerals written in the air with Leap motion sensor. In this paper, we also focus on proper handling of aerial input numerals to verify whether the numerals written in the air are helpful for authentication. We collect numerals 0 to 9 from five subjects, then apply three pre-processing to these data, learn and authenticate them by CNN (convolutional neural network) which is a method of machine learning. As a result of learning, an average authentication accuracy was 92.4%. This result suggests that numerals written in the air are possible to carry out personal authentication and it will be able to construct a better authentication system.

1. Introduction

With the progress of IoT, everything is going to be connected to the network. It will bring us a lot of benefits, however some security risks will be occurred by connecting network. To avoid such problems, It is indispensable to strengthen security more than now. As one of the security elements, We focused on personal authentication. As a current major personal authentication method, there are a method using card information and password or biometrics system using face, iris, vein and so on. Biometric authentication is possible with high accuracy, and it is difficult to duplicate authentication information. However, if biometric information is stolen, there is a big problems that it is difficult to change or the number of changes is limited. For method using card information and password, although there are some problems such as skimming and password leakage, it can be relatively easily dealt with by changing authentication information. Therefore, We considered that it is necessary to create a system that strengthens security by adding new elements to the existing authentication method using card information and password. Even if authentication information is stolen, this system can be easily recovered. As a method to realize this system, we propose a method input numerals in the air using Leap motion sensor and then perform numerals recognition and personal authentication from input handwriting.

Figure 1 : Leap motion sensor

For password authentication realization and further security enforcement, we proposed a method as following. First, numerals data are inputted using Leap motion. Then, these data are learned and identified with CNN which is one of deep learning methods. Moreover, we also assume that at the same time personal authentication will be carried out to enhance security. In this paper, we focus on personal authentication to verify whether the numerals written in the air have unique features between individual for authentication. We also focus on proper handling of aerial input numerals. Numerals written in the air have different features from numerals written in the paper. Therefore, we must try different pre-processing approaches to the data.

In the next subsection, we describe about related work to show the novelty in this paper.

* Shun Yamamoto, Email: c501837011@tokushima-u.ac.jp

1.1. Related work

First of all, this paper is an extended version of the work originally presented in 2018 IEEE International Conference on Internet of Things and Intelligence System (IOTAIS) [1]. In this work, we handled the data of three subjects measured by the Leap motion. Then we applied two pre-processings to these data. Finally they were learned by CNN. As a result, the average accuracy was 90.3%, False Rejection Rate(FRR) was 3.8% and False Acceptance Rate(FAR) was 5.9%. As an extension from this, in this paper, we carried out an addition of a pre-processing, increase of subjects, and change of the layer structure of CNN. By these extension, we could reach more effective way to handle the numerals written in the air, and we could confirm this layer structure change is useful for numerals which have fewer features between individual.

As a related work, there was a research conducted by Katagiri and Sugimura[2]. They used trajectory of writing in the air using a penlight and performed signature verification with DP matching on the data. As a result, FRR and FAR were both 3.6%. Moreover, they considered that it was difficult to duplicate the aerial input, which had individual features because the degree of freedom of input movement was large.

In the research by Hatanaka, et al. [3], they acquired signature data including finger identification information from 50 people by the Leap motion. Then, they also used DP matching to express a distance between features and carried out evaluation using the distance. Moreover, they experimented with or without screen feedback. When there were the feedback and no information of finger identification, the Equal Error Rate(EER) was 3.72%. When there was finger identification information, the EER was 2.67%. On the other hand, when there were no feedback and no finger identification information, the EER was 6.00%. Furthermore when there was information of finger identification, the EER was 5.11%. In response to this result, they considered that using of finger identification information could have a performance equivalent or superior to the related work [4]. However, in this research [3], subjects had to use several fingers when perform the measurement.

In the research by Nakanishi, et al. [5], they attempted to authentication of aerial handwritten signature only using information of fingertip coordinate without using the information of finger identification. They dealt with the time series data measured by the Leap motion and the subjects used only one finger for measurement. Then, another five people's forgery signature data were added to the kanji signature data for two kinds of people and the learning was done by CNN. As a result, the accuracy against two signatures were 97.0% and 95.9% respectively, FRR were 9.6% and 19.2%, and FAR were 0.8% and 0.1%. From these results, it indicated the possibility where the aerial input was useful even without using the information of finger identification. Assuming use in a real system, it is easy to use only one finger for users. Therefore, we attempt to do verification without the information of finger identification.

There is one another example using CNN and time series data [6]. In this paper, The feature extraction is carried out by CNN, as a conclusion, usefulness of using CNN for multivariate time series had been shown.

These researches indicate the usefulness of the aerial input for identification and authentication. These researches [2],[3],[4] dealt with hiragana or kanji in Japanese. We try to authenticate numerals in this paper. Numerals have relatively simpler shapes than hiragana or kanji. Therefore numerals data may not have big difference how to write. On the other hand, there was a result that the numerals written on a paper were identified in high accuracy (99.98%) [7].

Therefore, we try to verify that the numerals written in the air are useful for identification or authentication in this paper.

About numeral identification, there was our previous research [8]. In this research, we acquired 450 data from subjects by the Leap motion. We applied two pre-processings to these data, then these data were learned and classified by CNN. As a result, the average accuracy for identification was 93.4%. From this research, we could confirm the usefulness of numerals written in the air for identification. Hence, we try the personal authentication with the numerals written in the air as the next step in this paper.

2. Methods

In this paper, we measured numerals data using the Leap motion and then these data were learned and authenticated using CNN(Convolutional Neural Network). This flow is shown in Figure 2.

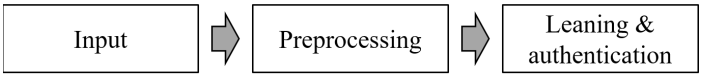


Figure 2 : Flow of the proposed method

When the Leap motion is used, a subject can write numerals with natural movement because it does not need special operation. Moreover, the introduction cost of the Leap motion is low and there was a result that characters which were measured by the Leap motion were useful for authentication. From these reasons, we use the Leap motion to collect data.

2.1. Input

The input data are measured by the Leap motion. Numerals 0 to 9 are the targets of measurement. These data consist of x, y, z coordinate values of the locus of the fingertip when subjects draw numerals. An example of the input data and its locus of x, y, z coordinate values are shown in Figure 3.

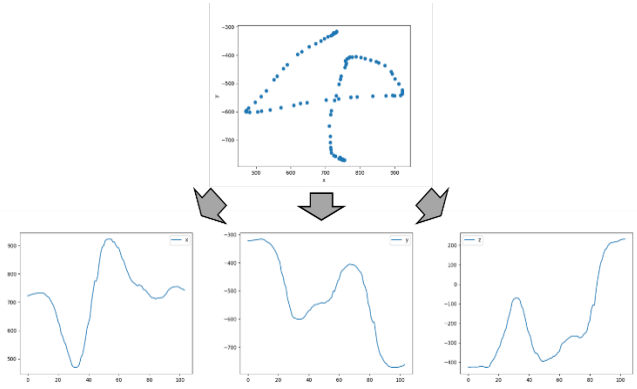


Figure 3 : Example of input data

2.2. Pre-processing

The data obtained directly from subjects are not suitable to learn in order to carry out the neural network learning. Therefore, we apply three pre-processings to raw measurement data then we reshape them to suitable. Of the three pre-processings, an outlier processing is the process newly added in this paper.

2.2.1 Adjustment of position to start writing

When we carry out the measurement, a inputting start coordinate values are not always same. From this factor, using such data directly for learning, they did not have good result for identification. To deal with this problem, we need to unify the writing start coordinate values. The i-th coordinate value before adjustment is defined as α_{before}, the i-th coordinate value after adjustment is defined as α_{after}, and the writing start coordinate value is defined as d. Moreover, we applied following formula to x, y, z coordinate values respectively for adjustment.

$$\alpha_{after}[i] = \alpha_{before}[i] - d \tag{1}$$

The data after application of this processing are shown in Figure 4. Through this adjustment, each starting point is adjusted to 0.

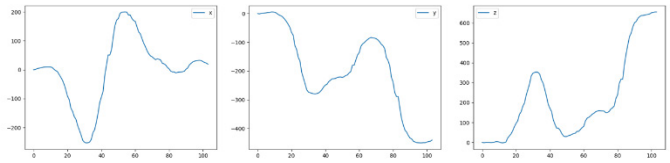


Figure 4 : Data after adjustment

2.2.2 Equalization of the data length

The measurement data have different data length for each measurement. To carry out learning by CNN using these data in later process, we have to unify length of the data. We set the longest data length in all measurement data as a length of the data after unified. Then, we use linear interpolation to unify the data length. The linear interpolation is a method to calculate a value corresponding to any coordinate which exists between two points. To carry out this processing, the following formulas are applied to i-th x, y, z coordinate values. In the following formulas, dst_size means data length after unification, $data_size$ means data length before unification, and t' means a value obtained by truncating the fractional part of t. The data after application of this processing are shown in Figure 5. In this paper, the number of data points was unified to 300.

$$scale = \frac{dst_size}{data_size} \tag{2}$$

$$t = \frac{i}{scale} \tag{3}$$

$$a = t - t' \tag{4}$$

$$\mathrm{x}' = \mathrm{x}_{[\mathrm{i}]}(1-a) + x_{[i+1]}a \tag{5}$$

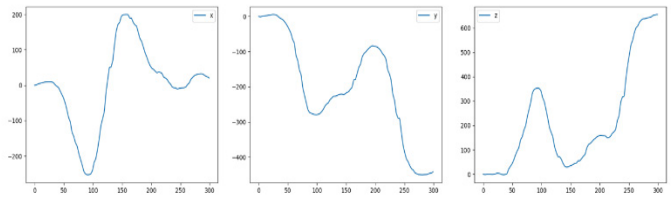


Figure 5 : Data after equalization using the linear interpolation

2.2.3 Outlier processing

When we carried out measurement, some outliers caused by the finger-sensor position relationship could occur. Therefore, it is necessary to remove such outliers. First, we use Interquartile range (IQR) to detect the outliers. We set upper limit and lower limit using IQR. Then, if any values exceed these limits, such values are detected as outliers. The IQR, upper limit, and lower limit can be calculated using the following formulas. 25percentile means 25 percent of data from the bottom and 75percentile means 75 percent of data from the bottom when data are arranged in ascending order.

$$IQR = 75percentile - 25percentile \tag{6}$$

$$Upper\ limit = 75percentile + 3.0 \times IQR \tag{7}$$

$$Lower\ limit = 25percentile - 3.0 \times IQR \tag{8}$$

Next, we remove the values which were detected by IQR. Finally we interpolate them by linear interpolation. These flow of processing is shown in Figure 6.

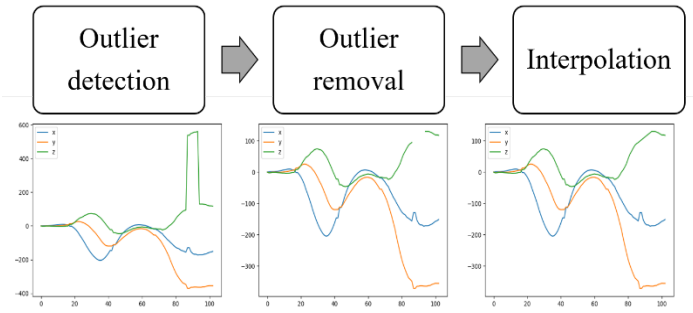


Figure 6 : Flow of outlier processing

By these processings, reduction of noises can be expected, we will be able to handle aerial input numerals better.

2.3. Learning & authentication

The preprocessed data are fed to CNN for authentication. The authentication (classification of subjects) with 2 classes separation(one subject or the others) is carried out.

2.4. Experiment

We measure the numerals 0 to 9 in order. Subjects carry out actual measurement after some practice of drawing. As shown in Figure 7, when we carry out the measurement, the Leap motion is placed on the desk and the subjects can see feedback on the screen.

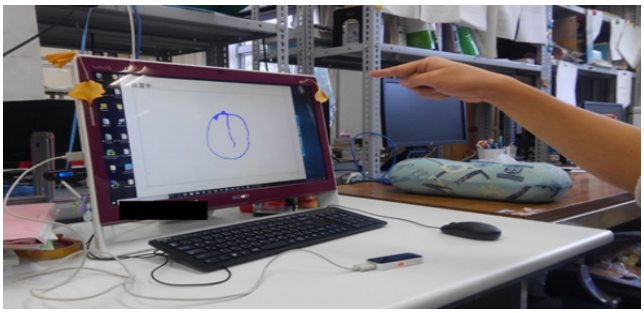


Figure 7 : Measurement scene

In the measurement, the subjects draw the numerals 0 to 9. The numerals 0 to 9 are defined as one set. 15 sets measurement a day is carried out. Measurement flow is shown in Figure 8.

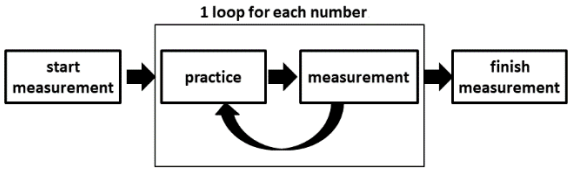


Figure 8 : Measurement flow

In this experiment , the subjects are four men and one woman. Measurement was carried out three days. These subjects are defined as A, B, C, D, and E. A target for authentication is the subject A. In the learning phase, a training data contains the two days data and a test data contains the rest one day data. Against these data, cross validation was carried out. The details of the data for validation are shown in Table 1.

Table 1 : Details of the data for cross validation

	Train	Test
Data1	Day1, Day2	Day3
Data2	Day1, Day3	Day2
Data3	Day2, Day3	Day1

Using these data, we carried out the learning and authentication by CNN. We used Adam as an optimization function and ReLU (rectified linear unit) as an activation function. Moreover, we set dropout rate of each layer 50%. The number of epochs was 300. The layer structure of CNN is shown in Figure 9.

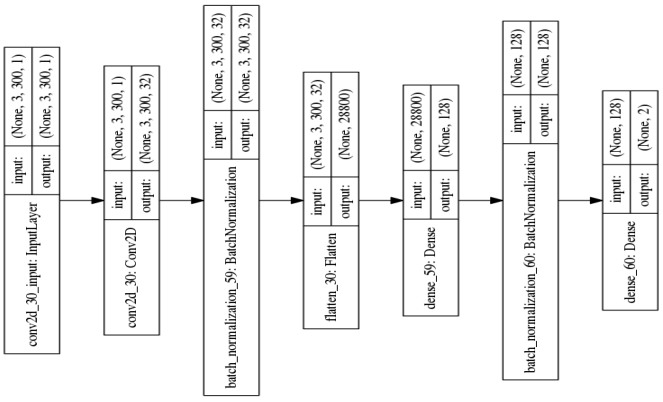


Figure 9 : Layer structure of CNN

As Figure 9 shows, the structure which was used in this experiment had just one convolution layer. Furthermore, the batch normalization and the early stopping were introduced to avoid the overfitting as much as possible. The kernel size was set to 3×2 to learn 3 channels(x, y and z coordinates) of input simultaneously. Moreover, when the convolution processing was carried out, we used a dilated convolution [9][10]. The dilated convolution can be carried out with making space for convolution. Using the dilated convolution, wider information can be processed. In the research by Luo, et al.[10], they referred dilated convolution is one of the methods to increase receptive field. Therefore, it can be expected it has the usefulness to authentication by numerals which have few feature differences between individual. This structure had fewer number of layers than common ones. There were two factors to this. First one was that the authentication target was numerals. Numerals have fewer features than other characters. The second factor was that the number of data was not sufficient for learning. From these factors, if there were more layers, excessive feature extraction could occur, and as the result, the overfitting may occur.

Table 2 : Experimental results 1

	Accuracy(%)		
	Data1	Data2	Data3
0	100.0	92.4	97.8
1	96.4	100.0	96.0
2	92.4	98.2	88.0
3	96.9	91.1	87.1
4	98.2	97.3	96.4
5	98.2	96.4	97.8
6	92.4	89.8	83.6
7	93.8	78.7	90.7
8	84.4	88.4	83.6
9	88.4	96.4	82.2

Table 3 : Experimental results 2

	Accuracy(%) (average)	FRR(%)	FAR(%)
0	96.7	0.0	5.4
1	97.4	0.5	3.7
2	92.9	1.5	10.4
3	91.7	0.2	13.6
4	97.3	1.5	3.0
5	97.5	1.0	3.2
6	88.6	2.5	16.5
7	87.7	0.7	19.8
8	85.5	5.2	19.0
9	89.0	0.7	17.5
average	92.4	1.4	11.2

3. Result

We carried out learning and authentication five times. The authentication results are shown in Tables 2 and 3. Moreover, we use FRR(False Rejection Rate) and FAR(False Acceptance Rate) as an evaluation criteria. FRR means a probability to reject a person who should be authenticated, and FAR means a probability to accept a person who should not be authenticated.

3.1. Discussion

We discuss about experimental results to verify the usefulness of the pre-processings and cause of misauthentication. From experimental results, we focus on numerals 5 and 8. These had the best and worst results. We considered that we could find out problems by comparing these numerals. First, comparison between subjects for data of "5" is shown in Figure 10.

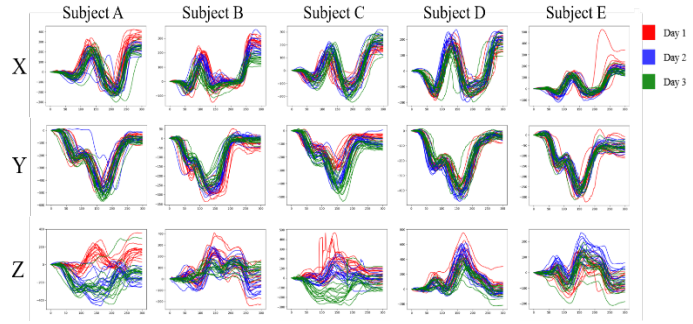


Figure 10 Comparison between subjects for data of "5"

The results for "5" were the best. From Figure 10, it can be seen that the x and y coordinate values have unique features between subjects. Even if the data had similar outlines, the scale of those data was different for each subject. However, in the z coordinate values, it can be seen that though some of data had unique features, data were sometimes unstable even within the same subject, and the outliers still remained. Therefore, we considered differences in the x and y coordinate values contributed to high accuracy. On the other hand, comparison between subjects for data of "8" which had the worst result is shown in Figure 11.

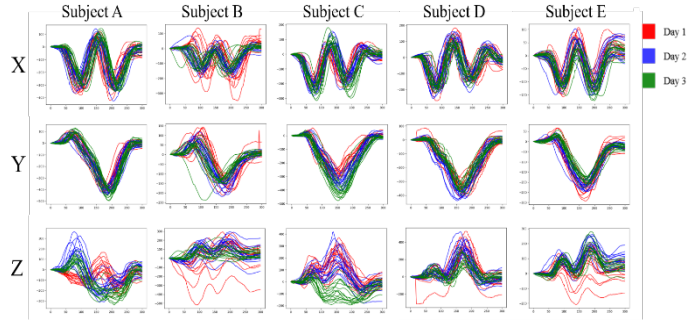


Figure 11: Comparison between subjects for data of "8"

From Figure 11, in the x and y coordinate values, it can be seen that the differences between subjects were relatively small compared to the case of "5". For the z coordinate values, they showed unique features for each subject and can be expected to be useful for authentication. However, some data trended differ even within the same subject. Thus, it can be considered the low accuracy was caused by such data. In the z coordinate values, such differences in data trended within the same subject can be seen not only in "8" but also in other numerals. Therefore, the misauthentication could be caused by the z coordinate values. However, if we removed the z coordinates, the authentication accuracy by numerals with relatively small difference between subjects in the x and y coordinate values bacame worse by reduction in subject's unique features. Accordingly, we will continue to use the z coordinates. However, we have to confirm about a uniformity of the data within the same subject to use these data appropriately. Hence, we will conduct an additional experiment for confirmation of the uniformity as a future plan.

Next, we discuss about the usefulness of the pre-processings. We used three pre-processings, and focus on the outlier processing among them. The outlier processing was the most recently introduced process in our research and its utility has not been tested yet. Moreover, the FAR was worse than before applying the outlier processing. This process may have lost the features of the data. When we carried out comparison of data before and after applying this processing, we found some problems. Examples of data that have a problem process are shown in Figure 12.

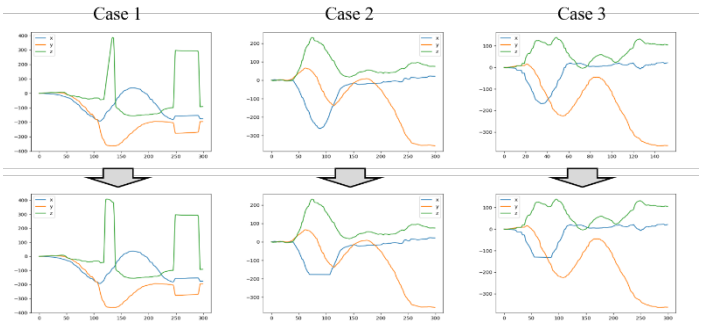


Figure 12 : Example of data that have a problem process

As shown in Figure 12, there were three cases. First, in the case 1, the outliers were not detected well although there were relatively long continuous outliers. If there were a lot of outliers, data distribution might be biased and outliers might not be detected. This distribution bias had a significant impact on the interquartile range. As a result, it can be considered this problem occurred.

Second, in the case 2, normal values were recognized as outliers. Therefore, a part of the data was deleted as outliers and an abrupt change occurred. Even if it looked like a natural change visually, it was regarded as outliers in case that the number of the data was small.

Third, in the case 3, the use of the linear interpolation for the outlier processing might be inappropriate. As shown in Figure 12, the linear interpolation sometimes caused unnatural interpolation.

However, the overall accuracy was improved by the introduction of the outlier processing because this processing was effective for some data. A successful example of outlier processing is shown in Figure 13.

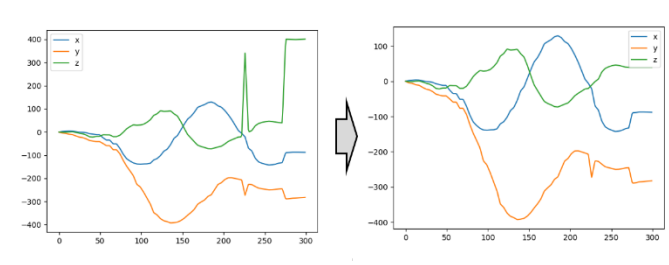


Figure 13 : Successful example of outlier processing

As shown in Figure 13, the outlier processing could reduce noises against some data. Noise generation is inevitable due to the nature of the experiment. Therefore, we need to solve the problems

mentioned in Figure 12. We will try to improve outlier processing by changing thresholds or trying another method as a future plan.

4. Conclusion

We proposed a method to carry out numeral identification and personal authentication by numerals inputted by the Leap motion sensor. In this paper, we focused on personal authentication and attempted discussion whether the aerial input numerals were useful for the authentication system. Moreover, we applied new one pre-processing and dilated convolution to numerals written in the air, verified these usefulness. We measured aerial input numerals data from five subjects by the Leap motion sensor and applied three pre-processings. And then, for these data, we carried out two classes separation(subject A or the others) by learning and authenticating using a Convolutional Neural Network with dilated convolution. As a result, average accuracy was 92.4%, FRR was 1.4%, and FAR was 11.2%. Compared to our previous paper[5], the overall accuracy has been improved. This result suggests that the aerial input numerals have enough features for authentication and the outlier processing and dilated convolution newly introduced this time are effective to aerial input numerals. Therefore we consider that we could show the usefulness of the aerial input numerals for authentication system. However, several problems, existence of biased z coordinate values in the same subject and problem of the outlier processing, were found by experiments and discussion. Against these problems, we will conduct additional experiment to collect more data, and then confirm trends of data. Moreover, we will try to carry out another method for outlier removal, find more effective way to handle aerial input numerals.

References

[1] Shun Yamamoto, Minoru Fukumi, Shin-ichi Ito, Momoyo Ito, "Authentication of Aerial Input Numerals by Leap Motion and CNN", Proc. of 2018 IEEE International Conference on Internet of Things and Intelligence System (IOTAIS), pp.189-193. https://doi.org/10.1109/IOTAIS.2018.8600847

[2] Masaji Katagiri, Toshiaki Sugimura, "Personal Authentication by Signatures in the Air with a Video Camera" PRUM 2001-125, pp.9-16(2001) in Japanese

[3] Issei Hatanaka, Masayuki Kashima, Kimionri Sato, Mutsumi Watanabe, "Study on Flexible Aerial Signature Individual Authentication System using the Finger Discrimination Information", Journal of the Institute of Image Information and Television, Vol70, No.6. pp.J125-J132(2016) in Japanese

[4] Kumiko Yasuda, Daigo Muramatsu, Satoshi Shirato, Takashi Matsumoto, "Visual-Based Online Signature Verification Using Features Extracted from Video", Journal of Network and Computer Applications, 33, 3, pp.333-341(2010). https://doi.org/10.1016/j.jnca.2009.12.010

[5] Kohei Nakanishi, Shin-ichi, Minoru Fukumi, Momoyo Ito, "Biometrics Authentication of Aerial Handwritten Signature Using a Convolutional Neural Network", Proc. of The 5th IIAE International Conference on Intelligent Systems and Image Processing 2017, 19-24. https://doi.org/10.12792/icisip2017.007.

[6] Zheng Yi., Liu Qi., Chen Enhong., Ge Yong., Zhao J.Leon. (2014) "Time Series Classification Using Multi-Channels Deep Convolutional Neural Networks." In: Li F., Li G., Hwang S., Yao B., Zhang Z. (eds) Web-Age Information Management. WAIM 2014. Lecture Notes in Computer Science, vol 8485. Springer, Cham. https://doi.org/10.1007/978-3-319-08010-9_33

[7] Sajjad. S. Ahranjany, Farbod. Razzazi and Mohammad. H. Ghassemian, "A very high accuracy handwritten character recognition system for Farsi/Arabic digits using Convolutional Neural Networks," 2010 IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA), Changsha, 2010, pp. 1585-1592. https://doi.org/10.1109/BICTA.2010.5645265.

[8] Shun Yamamoto, Minoru Fukumi, Shin-ichi Ito, Momoyo Ito, "Recognition of Aerial Input Numerals by Leap Motion and CNN", Proc. of SCIS and ISIS2018, pp.1177-1180. https://doi.org/10.1109/SCIS-ISIS.2018.00185.

[9] Fisher Yu, Vladlen Koltun, "Multi-Scale Context Aggregation by Dilated Convolutions", Proc. of International Conference on Learning Representation (ICLR2016). arXiv:1511.07122,

[10] Luo, Wenjie., Li, Yujia., Urtasun, Raquel., & Zemel, Richard. (2016). "Understanding the effective receptive field in deep convolutional neural networks". In Advances in neural information processing systems (pp. 4898-4906). arXiv:1701.04128

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 375-383 (2019

www.astesj.com

ASTESJ

ISSN: 2415-6698

Artificial Bee Colony-Optimized LSTM for Bitcoin Price Prediction

Andary Dadang Yuliyono, Abba Suganda Girsang*

Computer Science Department, BINUS Graduate Program-Master of Computer Science Bina Nusantara University, Jakarta, Indonesia 11480

ARTICLE INFO

Article history:
Received: 28 August, 2019
Accepted: 08 October, 2019
Online: 28 October, 2019

Keywords:
Hyperparameter
Long Short Term Memory
Artificial Bee Colony
TensorFlow
Deep Learning
Bitcoin

ABSTRACT

In recent years, deep learning has been widely used for time series prediction. Deep learning model that is most often used for time series prediction is LSTM. LSTM is widely used because of its excellence in remembering very long sequences. However, doing training on models that use LSTM requires a long time. Trying from one model to another model that use LSTM will take a very long time, thus a method is needed for optimizing hyperparameter to get a model with a small RMSE. This research proposed Artificial Bee Colony (ABC) as a method in optimizing hyperparameter for models that use LSTM. ABC is a metaheuristic method that mimics the behavior of bee colonies in foraging. Optimized hyperparameter in this research consisted of sliding window size, number of LSTM units, dropout rate, regularizer, regularizer rate, optimizer and learning rate. In this research the proposed method called as ABC-LSTM. Bitcoin prices historical data was used as the dataset for evaluating the prediction of the models. The best ABC-LSTM model resulted best RMSE of 189.61 compared to model that use LSTM without optimization resulted best RMSE of 236.17. This result showed that ABC-LSTM model outperformed models that use LSTM without optimization.

1. Introduction

Nowadays bitcoin has become the most popular cryptocurrency and has the largest capitalization value compared to others cryptocurrency. Besides being a digital currency, bitcoin has also become a trading instrument. The price of bitcoin has very high volatility. Even though it has high volatility and is full of risks and uncertainties, many people are interested in trading bitcoin. The high volatility of bitcoin price movements has encouraged many people to speculate on getting high capital gains. To assist bitcoin traders, a system that is capable to predict the price of bitcoin is needed, which intended to increase the potential to get large profits as well as to avoid potential losses

In recent years there have been many studies that use deep learning for predicting the price of bitcoin. Adebiyi et al. [1] and Guresen et al. [2] used Multi Layer Perceptron (MLP) while Jang and Lee [3] and McNally et al. [4] used Bayesian Neural Network. Deep learning model that is widely used for predicting bitcoin prices is Long Short-Term Memory Neural Network (LSTM) [4], [5]. LSTM is a special modification of Recurrent Neural Network (RNN). LSTM is widely used in time series forecasting activities because of its ability in remembering long sequential data.

The use of model that use LSTM for predicting bitcoin prices has a disadvantage, it requires a long training time [4] and a suitable hyperparameter combination is needed to get optimal results. There were several researches that tried to optimize hyperparameters in models that use LSTM using metaheuristics such as Particle Swarm Optimization (PSO) [6], Genetic Algorithm (GA) [7] and Ant Colony Optimization (ACO) [8]. The use of metaheuristic for optimizing hyperparameters on model that use LSTM was expected to get a model with small error.

In this research, a sliding window was conducted on the closing price of bitcoin and bitcoin volume traded from the previous days as a feature for price predictions of bitcoin in the next day. Using sliding window can improve the accuracy of bitcoin predictions [9]. To get the appropriate sliding window numbers, the sliding window was used as one of the optimized hyperparameter. This research used Artificial Bee Colony (ABC) as a method for optimizing hyperparameter for models that use LSTM for bitcoin price prediction.

ABC is a metaheuristic algorithm based on foraging behavior of bee colonies [10]. ABC had been used as an optimization algorithm in many task [11], such as training neural network [12]. Besides that ABC also used as a machine learning method, such as

*Abba Suganda Girsang, Email: agirsang@binus.edu

clustering [13]. ABC has a good balance in exploiting and exploring solution [14]. In this research ABC was proposed as an optimization method for optimizing hyperparameter.

The main contribution of this research is combining ABC and LSTM to optimize the hyperparameter that resulted a model with small RMSE. Optimized hyperparameter in this research consisted of sliding window size, number of LSTM units, dropout rate, regularizer, regularizer rate, optimizer and learning rate. Chung and Shin [7] combined GA and LSTM for optimizing hyperparameter consisted of sliding window size and number of LSTM units.

This paper is structured as follows. Section 2 contains the literature review, Section 3 contains the proposed method in this research, which is the combination of ABC and LSTM for optimizing hyperparameter, Section 4 contains an explanation of dataset, implementation, experiment results and comparison with benchmark models and Section 5 contains conclusion of this research.

2. Literature Review

2.1. Related Works

An old model that is often used for forecasting is the Auto Regressive Integrated Moving Average (ARIMA). Saxena et al. [15] used ARIMA model and compared its performance with LSTM model for predicting bitcoin prices. Bitcoin price prediction using LSTM outperformed ARIMA [15]. Similar to stock prices, bitcoin prices is often nonlinear and non-stationary, thus the application of the ARIMA model for long-term predictions is quite difficult [16].

In recent years, there are many researches tried to predict price movement by using deep learning. Deep learning is basically Artificial Neural Network (ANN) with input layers, several hidden layers and output layers which are usually referred to as Multi Layer Perceptron (MLP). Adebiyi et al. [1] used ANN for predicting stock prices.

Guresen et al. [2] used MLP, Dynamic Artificial Neural Network (DAN2) and hybrid neural network by using a generalized autoregressive conditional heteroscedasticity (GARCH) for predicting a stock market index.

Bayesian Optimized Recurrent Neural Network was used by McNally et al. [4] for predicting the classification of bitcoin prices movement that slightly outperformed the ARIMA model. McNally et al. [4] also used LSTM model with accuracy that exceeded the Bayesian Optimized RNN but the time needed by LSTM for training data was much longer than Bayesian Optimized RNN. This LSTM model had been proven effective for predicting bitcoin prices [4]. Jang and Lee [3] also used Bayesian Neural Network for predicting bitcoin prices based on bitcoin time series data and other blockchain information.

A model that use LSTM was also used by Alessandretti et al. [5] for predicting the cryptocurrency market where LSTM performed better for predictions with longer data days than models that use Gradient Boosting Decision Tree. LSTM for predicting stock price movements used in [17]–[19]. Nelson et al. [20] used LSTM to predict stock market price movements. Gao and Chai [19] used a stock prediction model that combines LSTM with PCA. Table 1 shows summary of related works. It describes dataset and method that was used and the experiment results.

Table 1: Summary of related works

Ref	Dataset	Method	Result
[15]	Bitcoin Price	LSTM	LSTM outperformed ARIMA
		ARIMA	
[4]	Bitcoin Price	Bayesian Optimized RNN	LSTM achieved highest classification accuracy while RNN achieved lowest RMSE
		LSTM	
		ARIMA	
[5]	Bitcoin Price	Single Regression XGBoost	Regression XGBoost worked best for short time window while LSTM worked best for long time window
		Multi Regression XGBoost	
		LSTM	
		Simple Moving Average	
[9]	Bitcoin Price, prices of crude oil, SSE, gold, VIX, FTSE100, global currencies USD/CNY, USD/JPY, USD/CHF	Support Vector Regression	LSTM outperformed all other model
		Linear Regression	
		Neural Network	
		LSTM	
[21]	Bitcoin Price	PSO-MLP-NARX	The model was able to predict accurately while passing all model validation tests

Huisu et al. [9] used the LSTM model rolling window in several input features and information about blockchain information. This model accurately predicted the price of bitcoin. LSTM was also used by Yunbeom and Hwang [22] by comparing its performance with several neural networks namely Deep Neural Network (DNN), basic RNN, RNN with LSTM cell, RNN with GRU cell and bidirectional RNN. Best performance in specifity, precision and accuracy achieved by DNN while best performance in sensitivity achieved by Bidirectional RNN.

Some metaheuristic algorithms were used to optimize model that use LSTM such as Genetic Algorithm, Ant Colony Optimization and Particle Swarm Optimization [6]–[8]. Chung and Shin [7] used a model that use LSTM that is optimized using Genetic Algorithm (GA) for predicting the stock market. In metaheuristics, the trial and error method is usually used as the basis for estimation in determining the time window size and LSTM architecture to be used in the research model. In [7], a systematic method was used to determine the time window size and LSTM topology for predicting the stock market. In deep learning, it is very difficult to determine the optimal architectural parameters. [7] showed that the model can be used as an effective tool to determine the optimal or near optimal model in deep learning.

Chhachhiya et al. [6] designed architecture of a model that use LSTM by using Particle Swarm Optimization (PSO). This hybrid model optimizes 3 parameters, namely activation function, learning rate and hidden neurons. The parameter that produced the smallest RMSE value is the most optimal model.

Sheikhan and Mohammadi [23] used GA-ACO + PSO-MLP for predicting time series data. Sun et al. [24] combined AdaBoost-LSTM Ensemble Learning for conducting a financial time series Forecasting. Indera et al. [21] used MLP which was optimized by using PSO for predicting the price of bitcoin.

2.2. Long Short Term Memory (LSTM)

LSTM is a type of Recurrent Neural Network (RNN) where modifications are made to the RNN architecture by adding a memory cell that can store information for a long period of time [25]. LSTM was proposed as a solution to overcome the vanishing gradient or exploding gradient problems that occur in RNN. This problem occurs when processing very long sequential data. This gradient problem caused RNN failing to capture long term dependencies [15]. Vanishing or exploding gradient problem will reduce the accuracy of a RNN model that affected the output of a prediction [26].

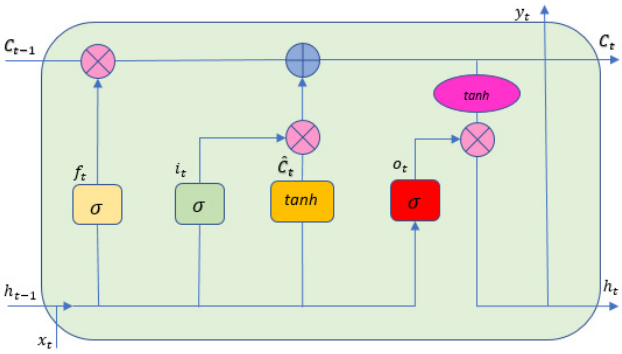


Figure 1: Architecture of LSTM

In a LSTM cell there is a memory cell and 3 gates, namely the input gate, the forget gate and the output gate. Figure 1 shows the architecture of LSTM and the processing of input data. The process of LSTM is carried out in the following stages [7]:

1. The value of an input can only be stored in the cell state only if it is permitted by the input gate. Calculation of the values at input gate and candidate inputs from the cell state is carried out using Eq. (1) and Eq. (2).

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (1)$$

Where i_t is the value of input gate, W_i is the weight for the input value at time to t, x_t is the input value at time to t, U_i is the weight for the output value from time to t-1, h_{t-1} is the output value from time to t-1 and b_i is the bias at the gate input and σ is the sigmoid function.

$$\tilde{C}_t = \tanh(W_c x_t + U_i h_{c-1} + b_c) \quad (2)$$

Where $\tilde{c}_t$ is the candidate cell state value, W_c is the weight for the input value in cell to c, x_t is the input value at time to t, U_i is the weight for the output value of cell to c-1, h_{c-1} is the value the output of cell to c-1 and b_c is bias in cell to c and *tanh* is a hyperbolic tangent function.

2. Then the forget gate value is calculated using Eq. (3).

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (3)$$

Where f_t is the forget gate value, W_f is the weight for the input value at time to t, x_t is the input value at time to t, U_f is the weight for the output value from time to t-1, h_{t-1} is the output value from time to t-1 and b_f is the bias on the forget gate and σ is a sigmoid function.

3. Furthermore the cell state memory is calculated using Eq. (4).

$$C_t = i_t * \tilde{C}_t + f_t * C_{t-1} \quad (1)$$

Where C_t is the memory cell state value, i_t is the value of the gate input, $\tilde{C}_t$ is the candidate memory cell state value, f_t is the forget gate value and C_{t-1} is the cell state memory value in the previous cell.

4. After generating a new cell state memory, the value of the gate output can be calculated using Eq. (5).

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (2)$$

Where o_t is the value of the gate output, W_o is the weight for the input value at time to t, x_t is the input value at time to t, U_o is the weight for the output value from time to t-1, h_{t-1} is the output value from time to t-1 and b_o is the bias at the gate output and σ is the sigmoid function.

5. The final output value is calculated using Eq. (6).

$$h_t = o_t * \tanh(C_t) \quad (3)$$

Where h_t is the final output, o_t is the gate output value, C_t is the new memory cell state value and the tan is the hyperbolic tangent function.

2.3. Artificial Bee Colony

Artificial Bee Colony algorithm was first proposed by Karaboga in 2005. It is a metaheuristic algorithm based on bee colonies. This algorithm mimics the intelligent behavior of bee colonies in foraging. In ABC, a colony of bees is divided into 3 types. Employee bee that is responsible for exploiting food sources, onlooker bee that is participating in exploiting food sources based on food information received from employee bee in waggle dance and scout bee that is tasked to find a new food source.

In ABC Algorithm, solutions of the optimization problem is described as a food source (nectar) [11]. And the quality of nectar describes the objective function of a solution. The number of food sources is the same as the number of employee bees, while the number of employee bees is the same as the number of onlooker bees. Bee behavior in foraging can be described as follows [11]:

1. At the initial stage of looking for food sources, bees start randomly exploring the area around the nest to get food sources.
2. After finding food sources, bees begin to become employee bees and start exploiting food sources found. After that. the employee bee will return to the nest by carrying nectar and unloading the nectar. After unloaded the nectar, the employee bee can immediately return to the food source or the bee can share information about the food source to other bees by doing Waggle Dance. The number of movements in dance shows the quality of nectar. If the nectar has run out, the employee bee will become a bee scout and start looking for other food sources randomly.
3. Onlooker bees waiting in the nest and choose the food source after watching waggle dance performed by employee bees.

Based on the behavior of bee colonies in foraging, the steps of the Artificial Bee Colony algorithm can be described as follows [12]:

1. Initialize the x_i solution using Eq. (7). Each solution is a vector with *D* dimensions.

$$x_{ij} = x_j^{min} + rand(0,1)(x_j^{max} - x_j^{min}) \quad (7)$$

Where $i = 1 \ldots SN$, $j = 1 \ldots D$. *SN* is the number of food sources (solution) and *D* is the number of optimized hyperparameters.

2. Evaluate all solutions
3. *Cycle* = 1. The optimal parameter search process is in cycles, $C = 1, 2 \ldots, MCN$.
4. Repeat step 5 to 12 until cycle = Maximum Cycle Number (MCN)
5. Create a new v_i solution for each employee bee from the x_i solution using Eq. (8) and evaluate the results. This new solution is a modification of the previous solution.

$$v_{ij} = x_{ij} + \emptyset_{ij} \ (x_{ij} - x_{kj}) \quad (8)$$

Where $k \in \{1,2, ..., SN\}$ and $j \in \{1,2, ..., D\}$ are randomly selected indexes.

6. Perform the selection process. If the cost function of the new solution is smaller than the previous solution, the solution will be updated to the new solution. The employee's bee will then remember the new, better solution.
7. Calculate the probability value p_i for the x_i solution using Eq. (9) and Eq. (10).

$$fitness_i = \begin{cases} 1/(1+f_i) & if\ f_i \geq 0 \\ 1 + abs(f_i) & if\ f_i < 0 \end{cases} \quad (9)$$

$$p_i = \frac{fitness_i}{\sum_{i=1}^{SN} fitness_i} \quad (10)$$

Where f_i is the value of the objective function solution *i*.

Where $fitness_i$ is the fitness value of the value of the objective function solution *i*. And SN is the number of solutions.

8. The onlooker bee evaluates the solution based on the information provided by the employee bee then selects which solution to follow based on the probability of p_i calculated using Eq. (10). Create a new v_i solution for each onlooker bee from the v_i solution selected based on the probability of p_i using Eq. (7) and evaluating the results. As with the bee employee, this solution is a modification of the selected x_i solution.
9. Perform the selection process. If the cost function of the new solution is smaller than the previous solution, the solution will be updated to the new solution. The employee bee will memorize the new solution.
10. If there is a bee when modifying a solution does not get a new solution that is better than the previous solution until certain limit, then the solution will be abandoned, and the bee becomes a scout bee and randomly searches new food sources using Eq. (7). The new solution will replace the old solution that was abandoned.
11. Save the best solution up to now
12. Cycle = cycle + 1

3. Proposed Method

3.1. The Concept

Optimized hyperparameter in this research consisted of sliding window size (h_1), number of LSTM units (h_2), dropout (h_3), regularizer (h_4), regularizer rate (h_5), optimizer (h_6) and learning rate (h_7). Each hyperparameter has a different range of values. The concept of optimizing hyperparameter in the model that use LSTM using ABC is to make these hyperparameters as dimension of problems, thus there are 7 dimensions of the problem. Each combination of the 7 dimensions of the problem is a solution that represents a model that use LSTM. Those solutions in the form of a combination of hyperparameters are described as genes in a chromosome in the genetic algorithm as can be seen in Figure 2. Those solutions will be trained and then predictions will be made. The prediction results of the model is measured using RMSE and considered as the fitness function of a solution on ABC.

h_1	h_2	h_3	h_4	h_5	h_6	h_7

Figure 2: dimension of hyperparameter (dimension of problem)

Next, those solutions will be optimized by changing one of the values of the 7 dimensions of the hyperparameter randomly and then conducting training and prediction using the modified solutions. If the modified solution is better than previous solution, the solution will be updated. The process continues until the maximum cycle numbers specified in the ABC parameters setting is reached.

3.2. Artificial Bee Colony – Long Short Term Memory (ABC-LSTM)

The hyperparameter optimization process on LSTM using the Artificial Bee Colony is carried out in following steps :

1. Randomly generate N initial solutions (food sources) using Eq. (7). In this case the solutions are hyperparameter combination for training LSTM.
2. Conducting LSTM training using the initial hyperparameter, evaluating and memorizing the lowest fitness function, in this case is the RMSE value of the prediction
3. Employee bee evaluates the RMSE produced by the initial solutions and tries to modify the hyperparameter combination to minimize RMSE using Eq. (8).
4. Onlooker bees choose the best hyperparameter combination based on probability on Eq. (9) and Eq. (10) and try to modify the hyperparameter combination to minimize RMSE
5. If the attempts to modify the hyperparameter exceeds the abandon limit value, the employee bee turns into a scout and creates a new hyperparameter combination using Eq. (7).
6. The process is repeated until it reaches the maximum number of cycles and produces a hyperparameter combination with the smallest RMSE value.

4. Experimental Results

4.1. Data Collection

Bitcoin prices dataset that was used in this research was downloaded from the coinmarketcap.com website. Historical data on downloaded bitcoin prices have daily intervals from December 27, 2013 to January 21, 2019. Bitcoin price data traded using US dollar currency units. Historical data on bitcoin prices consisted of the opening price (open), the highest price (high), the lowest price (low), closing price (close), the volume of traded bitcoin (volume) and market capitalization as can be seen in table 2.

Table 2: Illustration of bitcoin price dataset

Date	Open	High	Low	Close	Volume	Market Cap
27-Dec-13	763.28	777.51	713.6	735.07	46,862,700	8,955,394,564
28-Dec-13	737.98	747.06	705.35	727.83	32,505,800	8,869,918,644
29-Dec-13	728.05	748.61	714.44	745.05	19,011,300	9,082,103,621
30-Dec-13	741.35	766.6	740.24	756.13	20,707,700	9,217,167,990
31-Dec-13	760.32	760.58	738.17	754.01	20,897,300	9,191,325,349

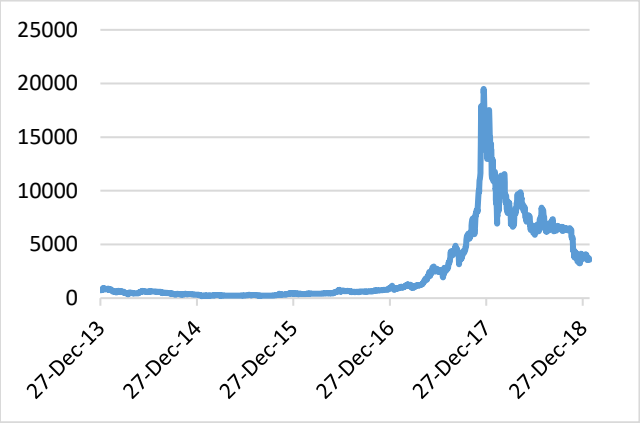


Figure 3: Bitcoin close price

Figure 3 shows the historical movement of bitcoin prices, while figure 4 shows the historical movement of bitcoin trading volume. Highest bitcoin price and volume of bitcoin trading throughout history occurred in December 2017. Bitcoin prices and bitcoin trading volume were used as features in this research.

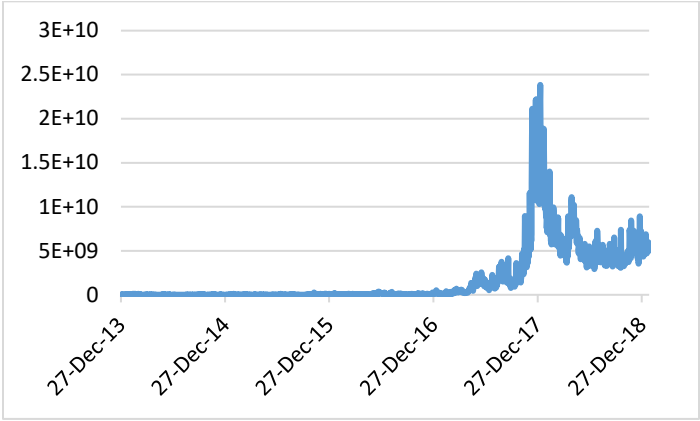


Figure 4: Bitcoin traded volume

4.2. Data Normalization

In this step the dataset value was converted into data with a scale from 0 to 1 using MinMaxScaler in a Scikit Learn library. The dataset after normalization can be seen in table 3. The dataset was divided into 3 sections consisted of 80% training dataset, 10% validation dataset and 10% test dataset. Illustration of bitcoin price dataset after normalized can be seen in table 3.

Table 3: Illustration of normalized bitcoin price dataset

Date	Open	High	Low	Close	Volume	Market Cap
27-Dec-13	0.03038	0.02846	0.02883	0.02883	0.00185	0.02009
28-Dec-13	0.02907	0.02693	0.02839	0.02846	0.00124	0.01983
29-Dec-13	0.02856	0.02701	0.02888	0.02935	0.00068	0.02048
30-Dec-13	0.02925	0.02791	0.03025	0.02992	0.00075	0.02090
31-Dec-13	0.03023	0.02761	0.03014	0.02981	0.00076	0.02082

4.3. Feature

Features that were used in this research consisted of the closing price of bitcoin and the volume of bitcoin traded as seen in table 4.

Table 4: Features

Feature	Description
P	Bitcoin close price
V	Bitcoin traded volume
Y	Predicted Bitcoin price

4.4. Hyperparameter

In this step, hyperparameters selection were determined, along with variations in values. This optimized hyperparameters had a significant effect on the performance of the LSTM model. This

research optimized 7 hyperparameters with varying values. Breuel [27] and Greff et al. [28] compared the LSTM performance using various hyperparameters, and concluded that one of the most important hyperparameter was learning rate, because it had direct impact to model performance. Selected hyperparameters and the range values in this research can be seen in table 5.

Table 5: List of optimized hyperparameter

No	Hyperparameter	Range Value	Interval
1	Sliding Window Size	40 - 70	5
2	Number of LSTM Units (neurons)	30 - 100	5
3	Dropout	0.3 - 0.5	0.01
4	Learning Rate	0.0001 - 0.01	0.0001
5	Regularizer	L1, L2, L1L2	-
6	Regularizer Rate	0.005 - 0.02	0.001
7	Optimizer	rmsprop, adam, nadam	-

4.5. Environment and Parameter Setting

These experiments were conducted on Google Colab using GPU accelerator. The models were created using TensorFlow. TensorFlow is an open source deep learning framework that is widely used for research and production. In this research were performed 30 experiments, where each experiment consistsed of approximately 210 models and each model used 512 batch size and 1.000 epochs.

ABC algorithm in this experiment used the parameter setting as in the table 6.

Table 6: ABC parameter setting

Parameter	Setting
Dimension	7
Solution Number	10
Population Size	20
Limit	7
Maximum Cycle Number	10

4.6. Evaluation

For evaluating the performance of all models with different hyperparameter combination that was performed in the ABC optimization, Root Mean Squared Error (RMSE) was used. RMSE is calculated using Eq. (11). RMSE is often used to evaluate the predictions of forecasting activities. The smaller the RMSE value, the more accurate the predictions.

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2} \qquad (11)$$

4.7. Experiment result obtained using ABC-LSTM

Figure 5 shows the convergence process of ABC-LSTM in obtaining a combination of hyperparameter that produced a prediction with the smallest RMSE value. From 30 experiments, where each experiment consisted of approximately 210 models, result 30 best models, the first ABC-LSTM best model was selected that produced the smallest RMSE value 183.342117450547. The first ABC-LSTM best model had hyperparameter as follows:

- Sliding window size = 60
- Number of LSTM neurons = 65
- Dropout = 0.31
- Learning Rate = 0.0091
- Regularizer = L2
- Regularizer Rate = 0.014
- Optimizer = RMSprop

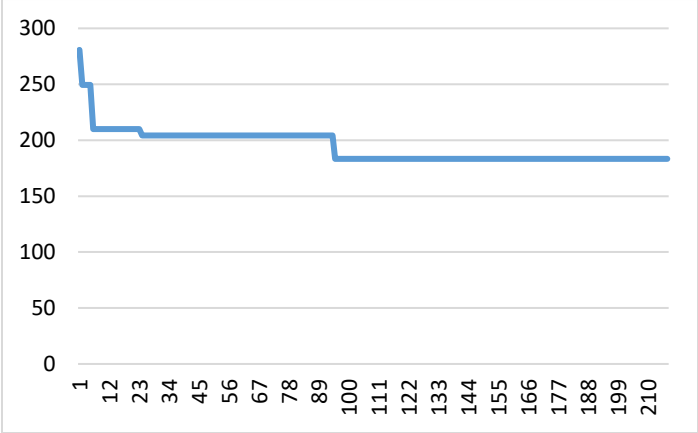


Figure 5: Convergence of the selected ABC-LSTM best model

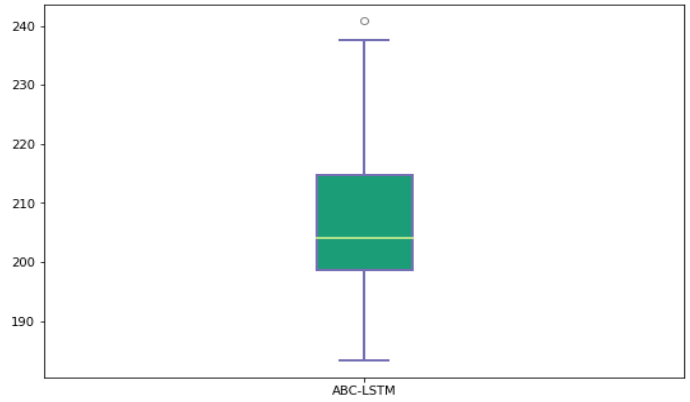


Figure 6: Boxplot of 30 ABC-LSTM best model performance

Figure 6 shows the boxplot of 30 ABC-LSTM best model from 30 experiments while table 7 shows the descriptive statistics of the RMSE of ABC-LSTM best model from each experiment. Smallest RMSE value obtained is 183.3421175 and biggest RMSE value obtained is 240.8371386.

4.8. Comparison with LSTM without optimization

After obtaining the best combination of hyperparameter, comparisons conducted to model that use LSTM without

optimization or using standard hyperparameter. For model without optimization there were 7 experiments with sliding windows of 5, 10, 20, 40, 60, 80 and 100, each with 60 runs. For the first ABC-LSTM best model also with 60 runs.

Table 7: 30 ABC-LSTM best model performance

Measure	ABC-LSTM
Mean	206.8833373
Standard Deviation	13.31792151
Best	183.3421175
Worst	240.8371386

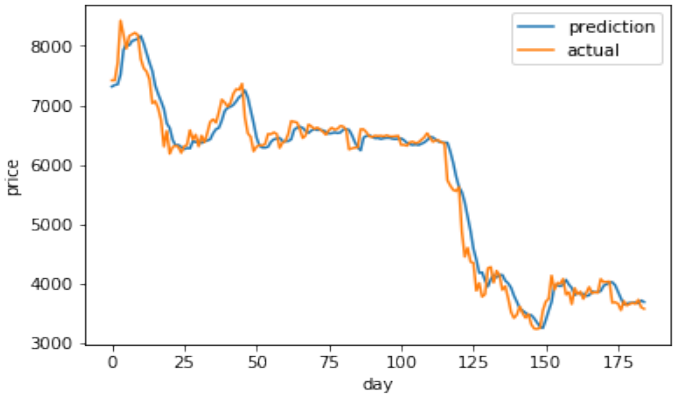


Figure 7: LSTM (5) prediction and actual comparison

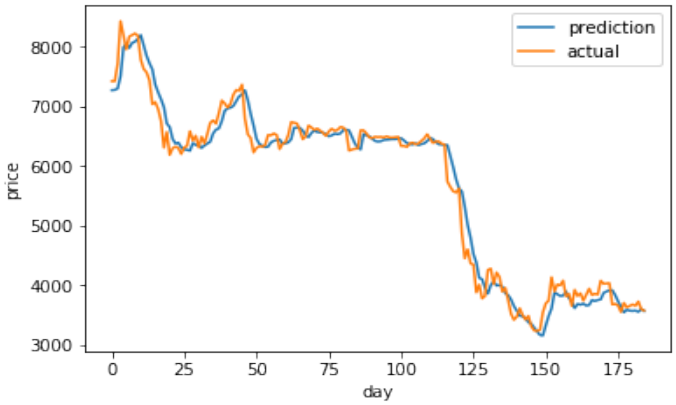


Figure 8: LSTM (10) prediction and actual comparison

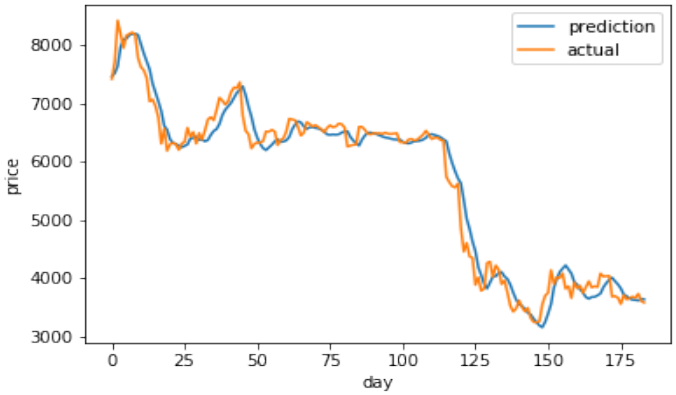


Figure 9: LSTM (20) prediction and actual comparison

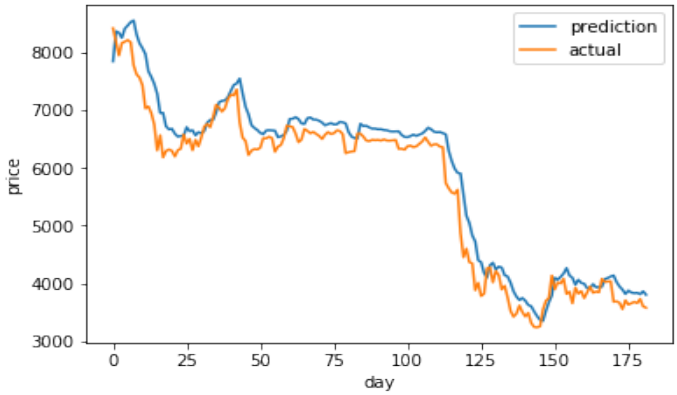


Figure 10: LSTM (40) prediction and actual comparison

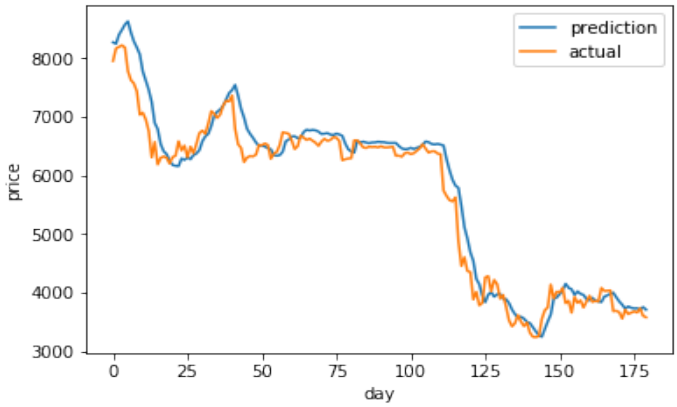


Figure 11: LSTM (60) prediction and actual comparison

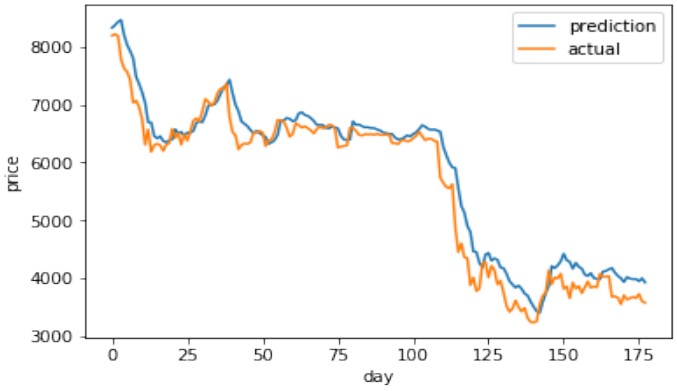


Figure 12: LSTM (80) prediction and actual comparison

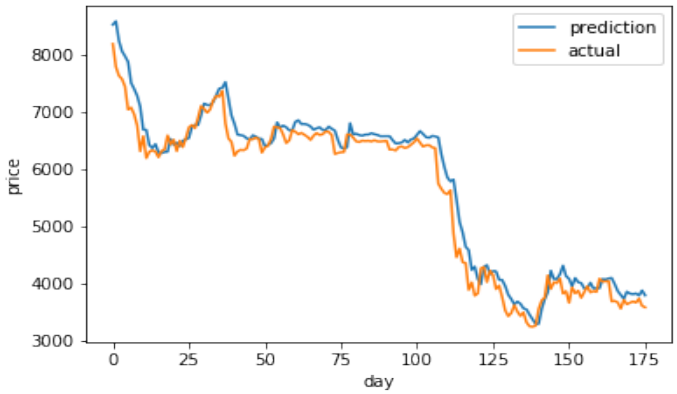


Figure 13: LSTM (100) prediction and actual comparison

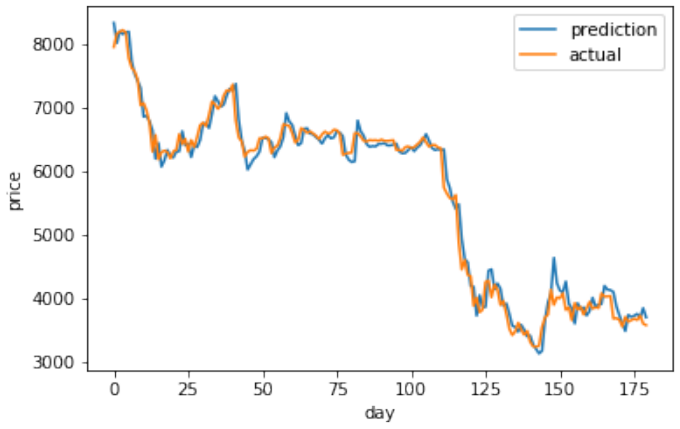


Figure 14: ABC-LSTM prediction and actual comparison

Figure 7-14 shows a comparison graph between the actual price and the prediction price of bitcoin using model that use LSTM without optimization, and the first best ABC-LSTM model. ABC-LSTM model shows most similarity between prediction and actual price.

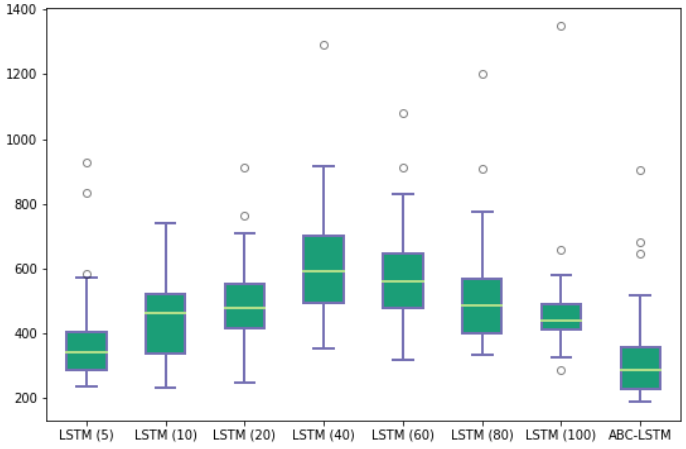


Figure 15: Boxplot of comparison with benchmark models

Table 8: Performance comparison

Model	Mean	Deviation Std	Best	Worst
LSTM (5)	368.62	125.94	236.17	929.33
LSTM (10)	437.88	114.34	233.66	740.96
LSTM (20)	488.13	124.33	246.52	912.74
LSTM (40)	611.44	162.23	351.73	1292.90
LSTM (60)	570.79	143.95	316.88	1080.47
LSTM (80)	503.91	145.46	332.63	1201.08
LSTM (100)	461.99	133.54	287.08	1347.78
ABC-LSTM	315.96	131.02	189.61	903.01

Figure 15 shows boxplot of comparison of model that use LSTM without optimization and the first ABC-LSTM best model. Prediction using the first ABC-LSTM best model produced the smallest RMSE value of 189.61. It also produced the lowest average RMSE value of 315.96. Performance of other models can be seen in table 8.

5. Conclusion

There were 2 features used in this study, bitcoin prices and bitcoin traded volume and 7 hyperparameters consisted of sliding window size, number of LSTM units, dropout, learning rate, regularizer, regularizer rate, and optimizer. From 30 ABC-LSTM experiments, the selected ABC-LSTM best model resulted RMSE of 183.34. The best ABC-LSTM model resulted best RMSE of 189.61 compared to models that use LSTM without optimization resulted best RMSE of 236.17. This result showed that ABC-LSTM model outperformed models that use LSTM without optimization.

Besides ABC, another metaheuristic that is very popular used for optimization is genetic algorithm (GA). GA is one of the evolutionary algorithms where best solutions will be used as parents. Those best solutions will be bred to produce a new solution that is better than the parent solutions. In the next research, a comparison between optimizing hyperparameter on models that use LSTM using ABC and GA will be performed.

References

[1] S. O. Adebiyi, A. A., Ayo, C. K., Adebiyi, M. O., & Otokiti, "Stock Price Prediction using Neural Network with Hybridized Market Indicators," J. Emerg. Trends Comput. Inf. Sci., vol. 3, no. 1, pp. 1–9, 2012.

[2] E. Guresen, G. Kayakutlu, and T. U. Daim, "Using artificial neural network models in stock market index prediction," Expert Syst. Appl., vol. 38, no. 8, pp. 10389–10397, 2011.

[3] H. Jang and J. Lee, "An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information," IEEE Access, vol. 6, no. 99, pp. 5427–5437, 2018.

[4] S. McNally, J. Roche, and S. Caton, "Predicting the Price of Bitcoin Using Machine Learning," Proc. - 26th Euromicro Int. Conf. Parallel, Distrib. Network-Based Process. PDP 2018, pp. 339–343, 2018.

[5] L. Alessandretti, A. ElBahrawy, L. M. Aiello, and A. Baronchelli, "Anticipating cryptocurrency prices using machine learning," Complexity, vol. 2018, pp. 1–16, 2018.

[6] D. Chhachhiya, A. Sharma, and M. Gupta, "Designing optimal architecture of recurrent neural network (LSTM) with particle swarm optimization technique specifically for educational dataset," Int. J. Inf. Technol., vol. 11, no. 1, pp. 159–163, 2019.

[7] H. Chung and K. Shin, "Genetic Algorithm-Optimized Long Short-Term Memory Network for Stock Market Prediction," Sustainability, vol. 10, no. 10, p. 3765, 2018.

[8] A. ElSaid, F. El Jamiy, J. Higgins, B. Wild, and T. Desell, "Using ant colony optimization to optimize long short-term memory recurrent neural networks," Proc. Genet. Evol. Comput. Conf. - GECCO '18, pp. 13–20, 2018.

[9] J. Huisu, J. Lee, and W. Lee, "Predicting Bitcoin Prices by Using Rolling Window LSTM model," ACM, 2018.

[10] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm," J. Glob. Optim., vol. 39, no. 3, pp. 459–471, 2007.

[11] B. Akay and D. Karaboga, "A modified Artificial Bee Colony algorithm for real-parameter optimization," Inf. Sci. (Ny)., vol. 192, pp. 120–142, 2012.

[12] D. Karaboga, B. Akay, and C. Ozturk, "Artificial Bee Colony (ABC) Optimization Algorithm for Training Feed-Forward Neural Networks," Modeling Decisions for Artificial Intelligence, pp. 318–329, 2007.

[13] D. Karaboga and C. Ozturk, "A novel clustering approach: Artificial Bee Colony (ABC) algorithm," Appl. Soft Comput. J., vol. 11, no. 1, pp. 652–657, 2011.

[14] M. Mernik, S. H. Liu, D. Karaboga, and M. Črepinšek, "On clarifying misconceptions when comparing variants of the Artificial Bee Colony Algorithm by offering a new implementation," Inf. Sci. (Ny)., vol. 291, no. C, pp. 115–127, 2015.

[15] A. Saxena, T. R. Sukumar, T. Nadu, and T. Nadu, "Predicting bitcoin price using lstm And Compare its predictability with arima model," Int. Journa Pure Appl. Math., vol. 119, no. 17, pp. 2591–2600, 2018.

[16] X. Zhan, Y. Li, R. Li, X. Gu, O. Habimana, and H. Wang, "Stock Price Prediction Using Time Convolution Long Short-Term Memory Network," KSEM, vol. 11061, pp. 461–468, 2018.
[17] S. Borovkova and I. Tsiamas, "An Ensemble of LSTM Neural Networks for High-Frequency Stock Market Classification," Journal of Forecasting, pp. 1–27, 2018.
[18] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," Eur. J. Oper. Res., vol. 270, no. 2, pp. 654–669, 2018.
[19] T. Gao and Y. Chai, "Improving Stock Closing Price Prediction Using Recurrent Neural Network and Technical Indicators," Neural Comput., vol. 30, no. 10, pp. 2833–2854, 2018.
[20] D. M. Q. Nelson, A. C. M. Pereira, and R. A. De Oliveira, "Stock market's price movement prediction with LSTM neural networks," in Proceedings of the International Joint Conference on Neural Networks, 2017, vol. 2017, pp. 1419–1426, 2017.
[21] N. I. Indera, I. M. Yassin, A. Zabidi, and Z. I. Rizman, "Non-linear Autoregressive with Exogeneous input (narx) bitcoin price prediction model using PSO-optimized parameters and moving average technical indicators," J. Fundam. Appl. Sci., vol. 9, no. 3S, p. 791, 2018.
[22] S. Yunbeom and Changha Hwang, "Predicting Bitcoin Market Trend with Deep Learning Models," Quant. Bio-Science, vol. 37, no. 1, pp. 65–71, 2018.
[23] M. Sheikhan and N. Mohammadi, "Time series prediction using PSO-optimized neural network and hybrid feature selection algorithm for IEEE load data," Neural Comput. Appl., vol. 23, no. 3–4, pp. 1185–1194, 2013.
[24] S. Sun, Y. Wei, and S. Wang, "AdaBoost-LSTM Ensemble Learning for Financial Time Series Forecasting," in International Conference on Computational Science, pp. 590–597, 2018.
[25] N. K. Manaswi, Deep Learning with Applications Using Python. Berkeley, CA: Apress, 2018.
[26] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: a deep learning approach for short-term traffic forecast," IET Intell. Transp. Syst., vol. 11, no. 2, pp. 68–75, 2017.
[27] T. M. Breuel, "Benchmarking of LSTM Networks," J. Cryst. Growth, vol. 285, no. 4, pp. 486–490, 2015.
[28] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A Search Space Odyssey," IEEE transactions on neural networks and learning systems, pp. 1–11, 2016.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 384-393 (2019

www.astesj.com

ASTESJ

ISSN: 2415-6698

Modelling and Control of The PV Power System Dedicated to The Greenhouse for Reduced Energy Consumption

Manel Jomaa UR-LAPER*, Mohamed Yassine Allani UR-LAPER, Abdelkader Mami UR-LAPER

Department of Physics, Faculty of Sciences of Tunis, University of Tunis El Manar, 2092 El Manar, Tunisia

ARTICLE INFO

Article history:
Received: 18 July, 2019
Accepted: 04 October, 2019
Online: 28 October, 2019

Keywords:
Greenhouse
Photovoltaic Generator
Permanent Magnet Synchronous Generator (PMSG)
Direct Control Torque (DTC)

ABSTRACT

A control of the greenhouse when it is connected to the photovoltaic generator is described in this paper. The greenhouse is modulated by differential equations while the temperature and humidity are regulated using the fuzzy logic controller. The photovoltaic system can supply maximum power for a specified voltage and output current. In addition, the fuzzy logic method is used to follow the maximum power point (MPPT) of a photovoltaic system. Our system consists of a solar module, a DC/DC converter with MPPT control, a DC/AC inverter and a three-phase induction motor. In addition, a direct torque control (DTC) of the induction machine which is utilized to control the ventilation flow rate. The state of control is displayed through the simulation results performed with Matlab / Simulink.

1. Introduction

Renewable energies provide an alternative option because they can replace conventional resources in different industrial applications. Renewable energy is a very important factor in greenhouse production. Indeed, a significant amount of energy consumed is required for heating in winter and ventilation in summer. To ensure high quality agricultural production, [1] guarantee an optimum regulation of greenhouse climate in an ecological and energy-saving way. The main purpose of a greenhouse is designed primarily to regulate its indoor climate to optimize crop production and protect plants [2]. The temperature and humidity are regarded both as important control elements [3]. A great deal of research on the control of greenhouse climates was presented in [4].

Some of the causes of the difficulties in controlling greenhouses are the complexity of a physical greenhouse model as well as the relation between the control actions and the processes that will be controlled. Heating, ventilation, humidification and dehumidification systems are designed to regulate the climate inside the greenhouse and to prolong production all year round.

Concerning the importance of indoor temperature and humidity, numerous regulation techniques such as the neural, fuzzy and adaptive regulation for optimizing the greenhouse environment have been used [5-7]. A PI technique was introduced in [8]. A neural network method was developed in [9] for the greenhouse system's complexity. [10, 11] was designed a regulation of the humidity and temperature in greenhouse by means of the fuzzy logic controller. Description of the Takagi-Sugeno method of air temperature control has been provided in [12]. [13] was based on predictive control.

For the intensive energy consumption in an agricultural greenhouse, it is important to find ways to reduce this consumption. In the following, a contribution to the management of energy consumption in an agricultural greenhouse through a photovoltaic ventilation system is proposed.

In recent years, a rapid development in the industrial uses of DTC technology has been demonstrated. The reasons for this are its rapid torque response, its easiness and robustness in response to changes in rotor parameters. As compared to the vector control diagram, DTC has delivered the same high performance dynamics with a simplified controller design as indicated in [14]. Therefore, we have chosen this method to control the ventilation motor.

The principal aim of this work is to contribute a high-tech control method and to ensure an appropriate greenhouse climate with lower energy consumption. First of all, a photovoltaic system will be developed. In a second step, we will present the greenhouse physical model. Then, we will introduce the fuzzy logic regulator used to control the indoor air temperature and relative humidity. A PV generator system is used to supply the ventilation motor via a DTC control using Matlab/Simulink software.

*Manel Jomaa, Bardo 2000 Tunis, manel.jomaa@enit.utm.tn

2. Photovoltaic modelling

Photovoltaic solar energy is an advantageous technology that has several benefits: the availability also in rural areas and the ideal conditions for an environmentally sustainable and efficient greenhouse.

2.1 Photovoltaic panel modelling

A suitable solution for the production of all the power required by the photovoltaic generator. A Matlab / Simulink modelling of the different parts of the system used. In general, the photovoltaic cell model is generally shown as a simplified circuit model equivalent, as illustrated by Figure 1, the equation is represented as follows [15].

$$I = I_{ph} - I_d - I_{sh} \quad (1)$$

I_{ph} : Photocurrent independent of V_{pv} (or R_s), it is proportional to the incident flux (recombination generation rate) and diffusion lengths of the carriers, so it is linearly dependent on solar radiation and impacted by the temperature according to the following equation:

$$I_{ph} = \left[I_{sc} + k_L \times \left(T_p - T_{ref} \right) \right] \times \frac{S}{S_r} \quad (2)$$

k_L: the short-circuit current coefficient / temperature

Where T_{ref} reference temperature, T_p is the actual temperature

S characterize the cell surface irradiation, (1000W/m2 is the nominal irradiation)

I_{sh}: Current flowing throughR_{sh}, if $R_{sh}$is very high, it becomes very low is independent of the voltage, and is given by:

$$I_{sh} = \left(\frac{v_{pv} + I \times R_s}{R_{sh}} \right) \quad (3)$$

$$I_d = Is \left[\exp\left[\frac{q_e \times V_{pv}}{A \times K \times T_p} \right] - 1 \right] \quad (4)$$

I_s Diode reverse saturation current, and it is given by:

$$I_s = I_{Rs} \left(\frac{T_p}{T_{ref}} \right)^{\left(\frac{3}{A_c} \right)} \times \exp\left[\frac{q_e \times E_{gap}}{A_c \times K} \left(\frac{1}{T_{ref}} - \frac{1}{T_p} \right) \right] \quad (5)$$

I_{Rs} : Saturation current, and it is given by:

$$I_{Rs} = \frac{I_{sc}}{\exp\left(\frac{q_e \times V_{oc}}{A_c \times K \times T_{ref}} \right) - 1} \quad (6)$$

Let us replace in equation (1) equations (3) and (4), the characteristic equation will become:

$$I = I_{ph} - I_s \left[\exp\left(\frac{q_e \times \left(V_{pv} + I \right) \times R_s}{A \times K \times T_p} \right) - 1 \right] - \left(\frac{v_{pv} + I \times R_s}{R_{sh}} \right) \quad (7)$$

With:

A_c : The cell ideal factor which can be determined by the recombination mechanisms in the Load zone of space

q_e : is the electron charge (1.6025 x 10~19c)

K: is the Boltzmann constant (1.38 10-23J/K).

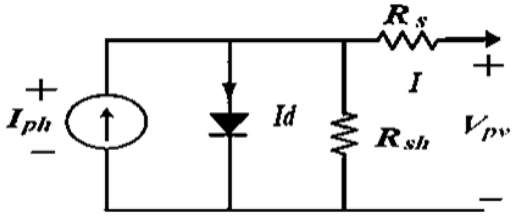


Figure 1: Electric model of the PV cell

The PV module configuration (Kaneka D-SA060) is associated in series (4 modules) and placed in parallel (7 modules). Table 1 summarises the photovoltaic generator parameters. [16]:

Table 1: Kaneka GSA060 value parameters

Parameters	Values
Maximum Power P_{max}	60W (+10/-5%)
Cells per Module	61
Voltage (P_{max})	67V
Current (P_{max})	0.91 A
OC voltage (V_{oc})	91.8 V
SC Current (I_{oc})	1.19 A
Shunt Resistance R_{sh}	254.8279 ohm
Series Resistance R_s	16.135 ohm

The system studied is composed by the PV generator and the DC-DC converter. DC-DC converter is controlled by the duty cycle generated by MPPT command based on fuzzy method. The variation of solar radiation and outdoor temperature that was measured on March 11th is shown in Figures 2 and 3.

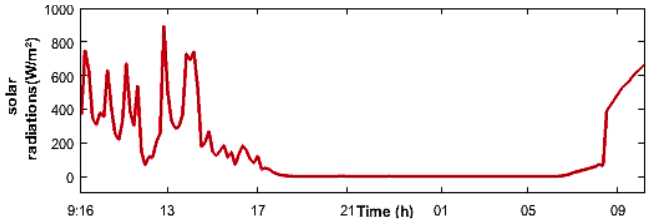


Figure 2: The radiation profile (11/03)

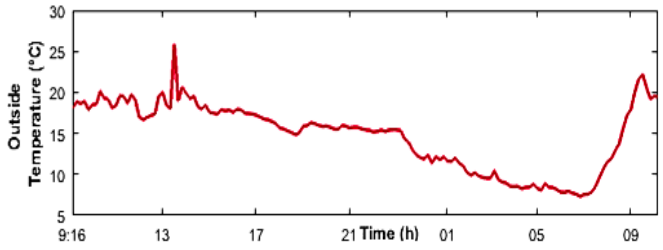


Figure 3: The profile of the outside temperature (11/03)

2.2 DC-DC boost converter modelling

The Boost converter is utilized in this task to augment the output voltage in relation to input voltage. Similarly, we also use it follow PPM tracking from the proposed control [17].

$$\begin{cases} v_0 = \left(\dfrac{1}{1-\alpha_{boost}}\right) v_{in} \\ I_0 = \left(1-\alpha_{boost}\right) I_L \end{cases}$$

With $0 < \alpha_{boost} < 1$

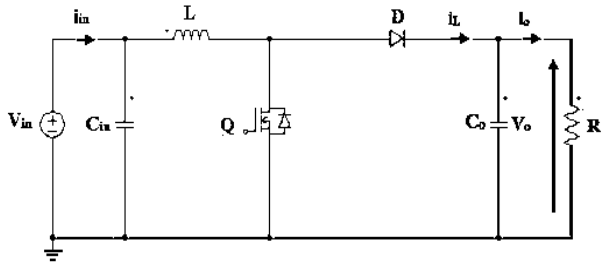


Figure 4: The electric model of DC-DC boost converter

2.3 Maximum Power Tracking Method

To control a Boost converter, we used the Fuzzy Logic method as this method is very practical and requires fewer measured parameters.

➢ *The Fuzzy Logic controller*

This command is designed to track the points of the maximum power. This command has the advantage of being robust. There are three steps considered to define the fuzzy controller concept: the fuzzification, the inference rules and the defuzzification [18], [19].

- The Fuzzification

This phase converts a physical variable into a linguistic variable. For every physical step, there is a degree of membership of a linguistic variable.

- The Rules of inference

This rule determines the controller's output signal in relation with the signals of input: those are indicated as "IF THEN". In fuzzy rules, operators "AND", "OR" intervene. For each rule, which is inserted by "THEN", the conclusion establishes a link between the membership factors of a premise to the membership function of the output variables.

- The Defuzzification

This step converts fuzzy output ensembles into physical variables suitable for such a process. Numerous defuzzification techniques exist, "centre of gravity" method is selected and described in the following equations:

$$x_{out} = \frac{\int x\mu(x)dx}{\int \mu(x)dx} \quad (10)$$

x_{out} : Output value. μ: Membership degree

- The fuzzification

In this step, we aim at designing a fuzzy controller using Perturb and Observe method.

Input variables are:

$$E_v(k) = V_p(k) - V_p(k-1) \quad (11)$$

$$E_P(k) = P_p(k) - P_p(k-1) \quad (12)$$

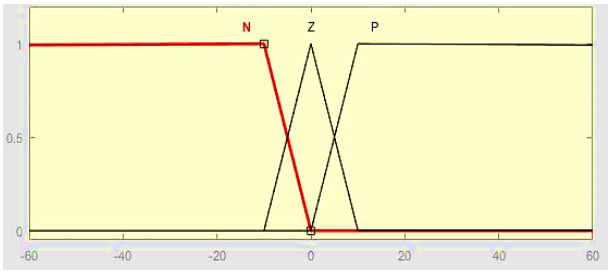


Figure 5: linguistic variable Ev (k)

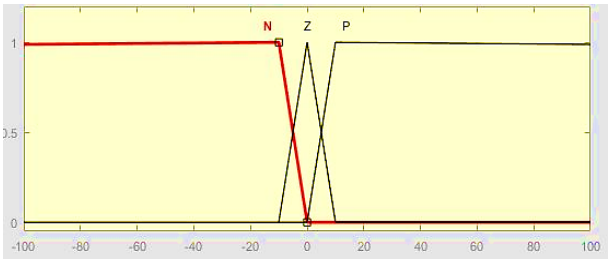


Figure 6: linguistic variable Ep (k)

- The inference Rules

The inference rules are described below:

Table 2: Rules of inference

Rules	Ep	Ev	D
1	p	P	z
2	p	N	Gp
3	p	Z	p
4	Z	Z	P
5	Z	N	P
6	z	P	p

7	N	P	Gp
8	N	N	z
9	N	Z	p

- Defuzzification

In Defuzzification we select the "centre of gravity" method that uses a single output variable called duty cycle D.

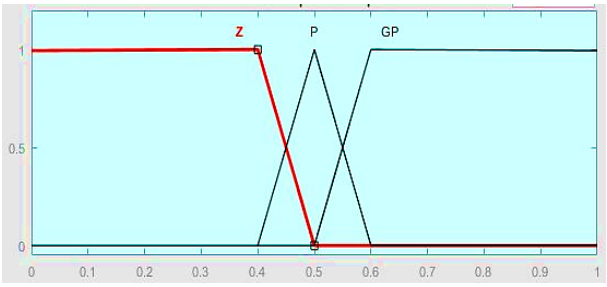


Figure 7: The Output of linguistic variable D (k)

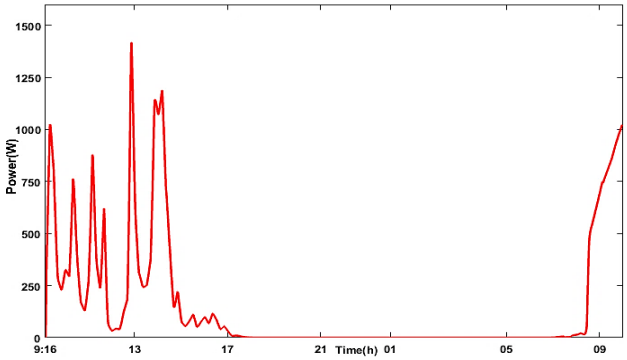


Figure 8: Output power supplied by the PV generator

3. Greenhouse Climate Model and Control

3.1. Model of the greenhouse

Inputs and outputs of a suggested model as shown in the next diagram. Input variables: MV_{α} the ventilation system, MV_{w} the heating system and MV_{fog} the fogging system, and the measured perturbations of wind speed, solar radiation, outdoor humidity and temperature [20].

The outputs variables: T_{int} indoor air temperature, H_{int} the indoor relative humidity.

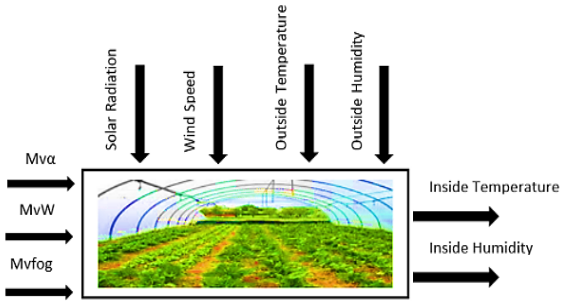


Figure 9: Model of the Greenhouse climatic [20]

3.2. Proposed model

- Heat Balance

In this section, we introduce the dynamic model of air temperature and thermal mass temperature inside greenhouse through the use of equations (1) and (9) [21].

$$v_i \rho c_p \frac{dT_i}{dt} = Q_{ss} - Q_{cc} + Q_{mm} - C_{sat}(Q_{ee} + Q_{nn}) - Q_v + W_h \quad (1)$$

Knowing that:

$$Q_{ss} = A_i \tau S_0 \quad (2)$$

Equation (2) is the solar energy delivered to the air volume.

$$Q_{cc} = A_i A_c \left(T_i - T_o\right) \quad (3)$$

Equation (3) is the energy transfer through conduction and convection processes.

$$Q_{mm} = A_i h_m \left(T_m - T_i\right) \quad (4)$$

Equation (4) is the energy exchange between the thermal mass and the inside air.

$$Q_{ee} = \lambda E_c \quad (5)$$

Equation (5) is the energy loss from culture evapotranspiration.

$$Q_{nn} = \lambda fog \quad (6)$$

Equation (6) is the energy loss by the nebulization.

$$Q_v = \rho c_p G(T_i - T_o) \quad (7)$$

Equation (7) is the energy exchange from window ventilation.

$$W_h = \frac{MV_w}{100} W_{\max} \quad (8)$$

The energy supplied by the heating system is provided by equation (8).

In addition, we describe in (9) the temperature of the thermal mass inside the greenhouse:

$$A_i C_m \frac{dT_m}{dt} = Q_{sm} - Q_{mm} - Q_{ff} \quad (9)$$

Knowing that:

$$Q_{sm} = \alpha_m Q_s \quad (10)$$

Equation (10) is the energy absorbed by the thermal mass by day.

$$Q_{ff} = A_i k_a \left(\frac{T_m - T_{ref}}{Z_{ref}} \right) \tag{11}$$

Equation (11) is the energy loss through the ground.

- Humidity Balance

Equation (12) determine the model of the humidity indoor the greenhouse [21].

$$\rho V_i \frac{dX_i}{dt} = F_{vv} + C_{sat}\left(E_c + fogg\right) \tag{12}$$

F_{vv} is the water rate in the air change flow determined by (13):

$$F_{vv} = \rho G\left(X_o - X_i\right) \tag{13}$$

E_c is the crop evapotranspiration determined by (14):

$$E_c = \frac{A_i\left(\Delta R_n + 2L\rho c_p D_i gwb\right)}{\left(\Delta + \gamma\left(1 + \left(\frac{gwb}{gws}\right)\right)\right)\lambda} \tag{14}$$

$fogg$ is the water rate of the fog system explained by (15):

$$fogg = \left(\frac{MV_{fog}}{100}\right) fog_{\max} \tag{15}$$

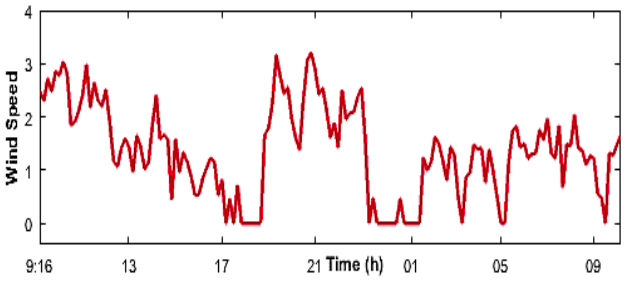


Figure 10: Wind speed (11/03)

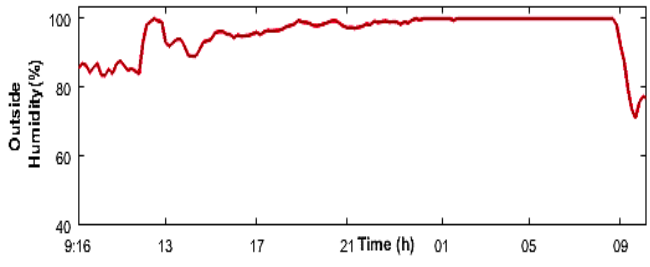


Figure 11: The outside Humidity (11/03)

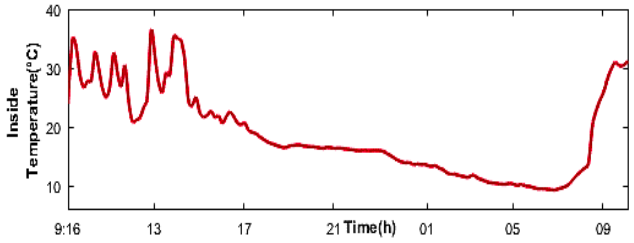


Figure 12: The results of the air temperature indoor greenhouse (11/03)

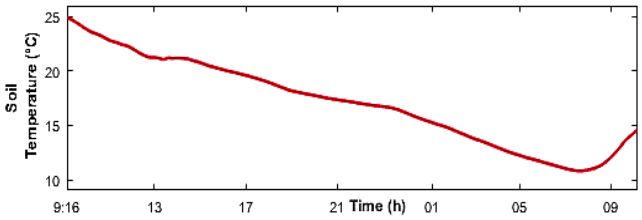


Figure 13: The results of the soil temperature indoor greenhouse (11/03)

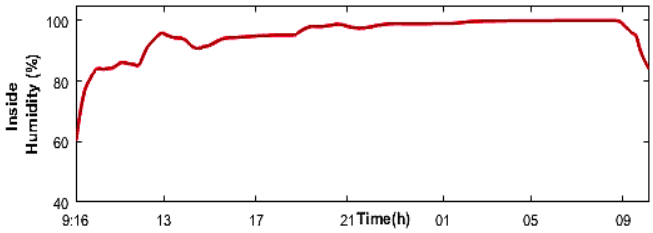


Figure 14: The results of the humidity indoor greenhouse (11/03)

Figure 10 illustrates the wind speed as measured on March 11th, while Figure 11 shows the outdoor humidity as measured on the same date.

Figures 12, 13 and 14 present simulation results for indoor air temperature, inside soil temperature (°C) and humidity (%).

The same greenhouse climate conditions as detailed above in [22] were utilized. That is situated in C.R.T.EN in Tunisia. Greenhouse axes are aligned in parallel with the East-West orientation. Its dimensions are 12.5 metres long, 8 metres wide and 3 metres high.

3.3. Fuzzy logic controller

Fuzzy logic controller is a sturdy methodology, and its theoretical method of development has been viewed as an interesting solution for a number of different engineering problems. This is thanks to its remarkable properties such as ruggedness, simplicity and flexibility. The fuzzy command is chosen since it is well suitable for multivariable non-linear systems [23, 24]. The methodology is based on three steps: Fuzzification, inference rules and defuzzification [25-27].

The principal purpose of the system is to control the climate inside greenhouse and to enhance its performance.

The fuzzy controller input variables are as follows:

Where:

$\Delta T \text{ int} = T_{desired} - T_{\text{int}} \in \left(Nb; Nm; Z; Pm; Pb\right)$

Outside temperature $\in$ (Pm; Z; Pb)

$\Delta H \text{ int} = H_{desired} - H_{\text{int}} \in (Nb; Nm; Z; Pm; Pb)$

Outside humidity $\in$ (Pm; Z; Pb)

Where:

$$\begin{cases} Nb : Negative_big \\ Nm : Negative_medium \\ Z : Zero \\ Pm : Positive_medium \\ Pb : Positive_big \end{cases}$$

$$\begin{cases} Pm : Positive_medium \\ Z : Zero \\ Pb : Positive_big \end{cases}$$

There are three output variables: heating rate (Heating), ventilation rate (Ventilation) and humidification rate (Humidification).

(Heating, ventilation) (Zero, medium, high)

(Humidifier) (Zero, Medium, High)

The fuzzy inference method is the Mamdani or the defuzzification technique used is based on the centre of gravity method.

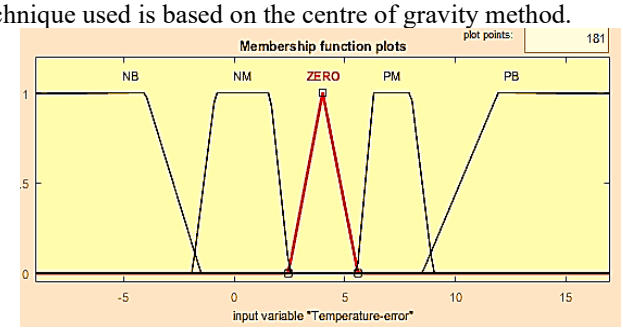


Figure 15: Membership functions of temperature error ($\Delta T \text{ int}$)

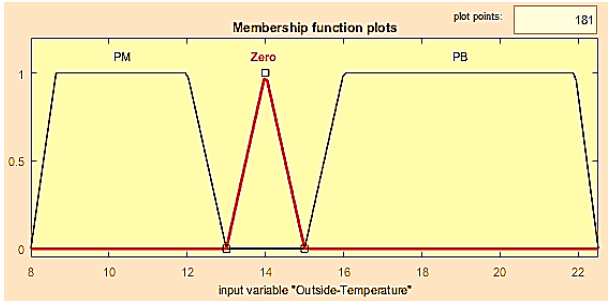


Figure 16: Membership functions of Outside Temperature

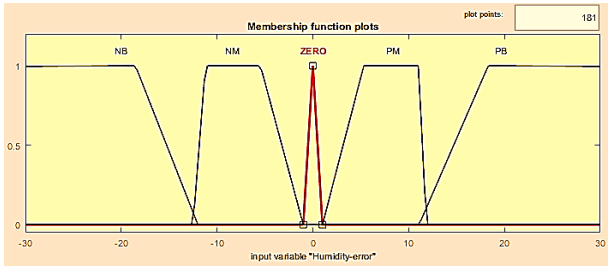


Figure 17: Membership functions of humidity error ($\Delta H \text{ int}$)

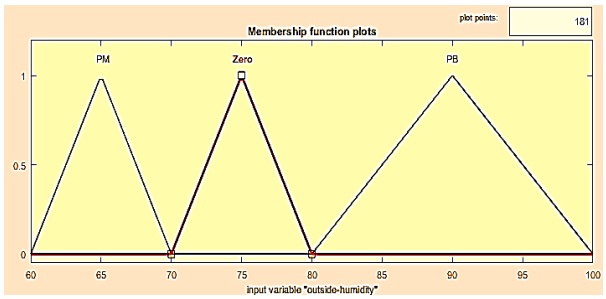


Figure 18: Membership functions of the Outside Humidity

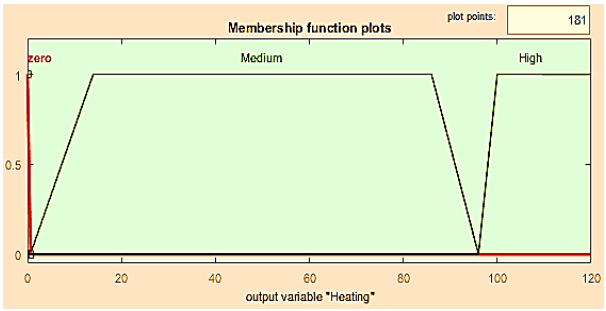


Figure 19: Membership functions of the heating rate

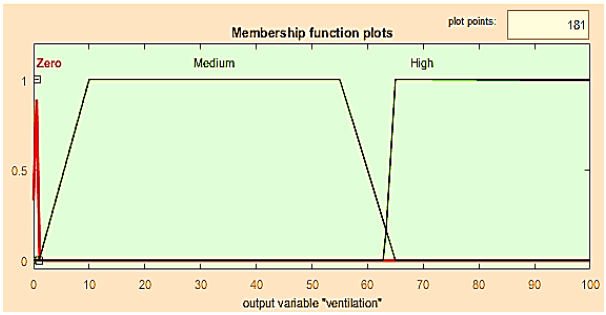


Figure 20: Membership functions of the ventilation rate

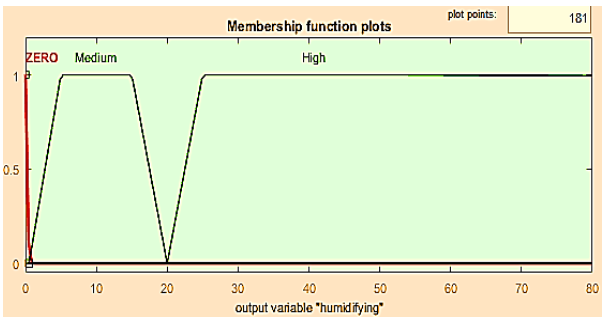


Figure 21: Membership functions of humidifying rate

Figures 15, 16, 17 and 18 illustrate the input variables of fuzzy logic regulator.

Figures 19, 20 and 21 present the output variables of fuzzy logic regulator.

4. DTC strategy

The ventilation technique currently used is based on the on-off control [28, 29]. The major disadvantage is high energy

consumption. Indeed, at the beginning of the ventilator motor, a large quantity of energy is required to reach maximum speed to reduce the temperature at the desired value. Therefore, using variable speed ventilation rather than the on-off control is more energy efficient.

In this work, a photovoltaic energy ventilation system is proposed, using the DTC direct torque control to regulate the ventilation motor as shown in Figure 22.

We used the DTC to regulate the motor that causes the ventilation. The greenhouse studied is equipped with a ventilation system based on a three-phase asynchronous motor with nominal power (705w)

Both the torque and flow technique can be changed rapidly by adjusting their settings. It also has high dynamic performance, a smaller switching frequency and consequently lower losses. This section focuses on a presentation about the photovoltaic ventilation system. This consists of a three-phase asynchronous motor with a fan as load torque. Such a load can be explained as a quadratic function [30] of motor speed as described in equation (1).

$$T_m = k \times w^2_m \tag{1}$$

With:

T_m : Mechanical torque

w_m : Motor speed

K is the constant that can be calculated by the equation (2).

$$P_{m_2 n} = T_{m_2 n} \times w_{m_2 n} \tag{2}$$

Where:

$P_{m_2 n}$: Mechanical power

$T_{m_2 n}$: Mechanical torque

$w_{m_2 n}$: The speed of the motor in its nominal state.

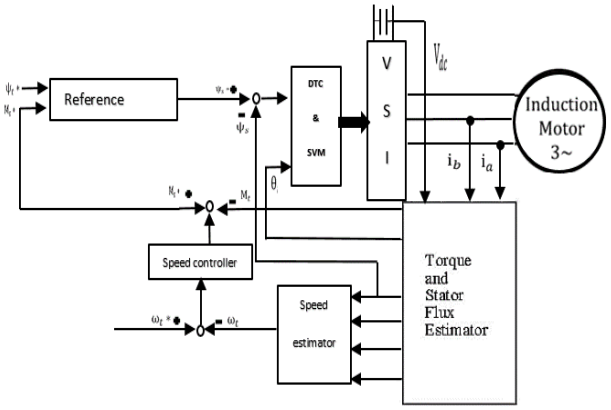


Figure 22: The block diagram of DTC for three-phase (IM) drives. SVM: space vector modulation

The DTC method is a technique that is utilized in variable frequency drives in order to regulate both the torque and the speed of three-phase electric motors.

This consists of determining a magnetic flux and motor torque estimate from the measured motor voltage and current. The basic diagrams of DTC Diagram for three-phase asynchronous motors with a speed regulation circuit are displayed in Figure 22.

The torque value is calculated by a proportional integral controller whose equation is expressed in equation (3).

$$T_{ref} = K_P \left[\Delta W + \frac{1}{T_i} \int \Delta W . dt \right] \tag{3}$$

ΔW is used to determine the difference between the speed reference and the motor's actual speed. The ventilation system speed setpoint is coupled to an air flow rate by using the ventilator law through a linear formula specified by equation (4).

$$V = W \times D^3 \tag{4}$$

With:

V: the requisite air flow rate; W: motor speed; D: fan diameter

5. Simulation and discussion

Figure 23 presents the results of the simulation of the fuzzy regulation of the greenhouse air temperature. There are two reference temperatures used as thermal conditions for optimal plant growth: 28°C during the day, 15°C in the night. Simulation results, illustrated in Figure 24, indicate that the humidity is maintained at about 70% of the desired value. Heating, ventilation and humidification rates are shown in Figures 25, 26 and 27 as well.

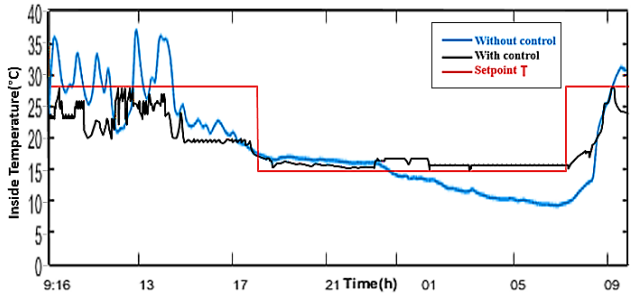


Figure 23: Results of greenhouse temperature with the controller Fuzzy Logic (11/03)

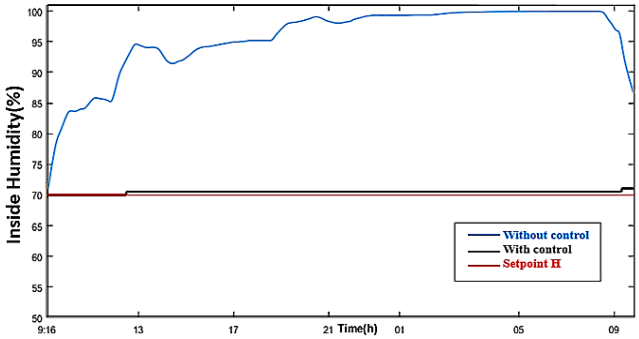


Figure 24: Results of greenhouse humidity with the controller Fuzzy Logic (11/03)

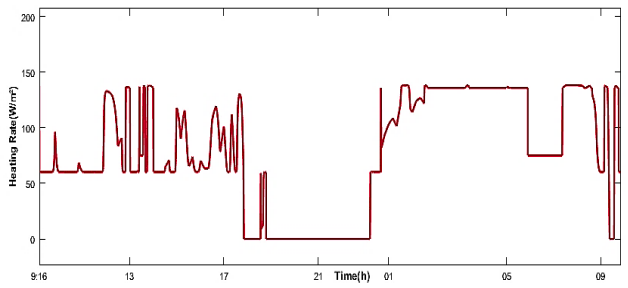


Figure 25: MVw [%]

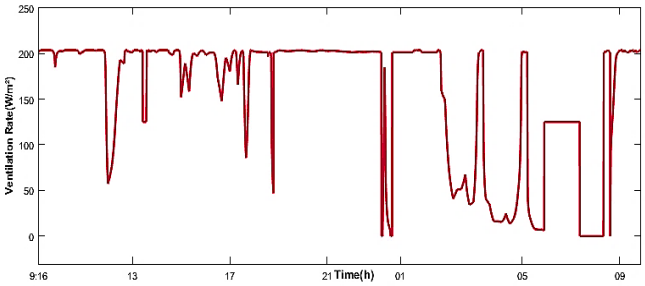


Figure 26: MVα [%]

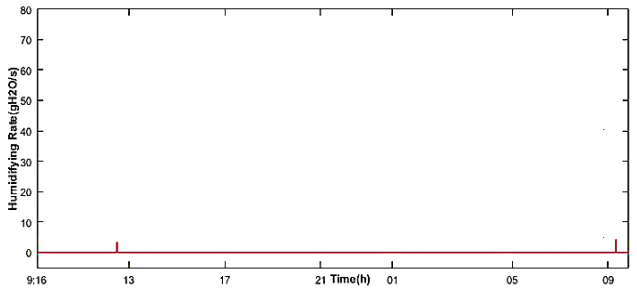


Figure 27: MVfog [%]

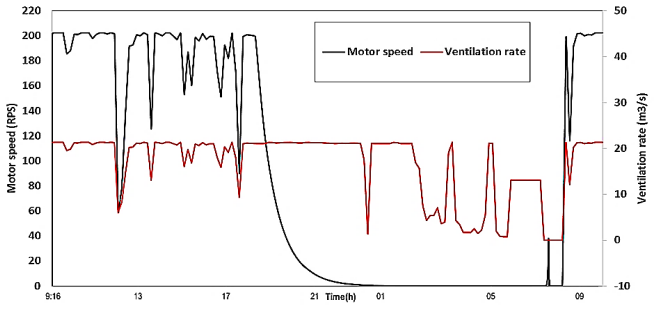


Figure 28: The results of ventilation rate and the motor speed

The principal contribution in this document consist on the development of photovoltaic energy to reduce the overloading of the electricity grid. On the basis of this compromise, Variable speed fan is introduced utilizing the (DTC) to support robust control of ventilation rate. Figure 28 illustrates the tracking of the motor speed. The results of the simulation demonstrate that the motor speed responds to the air flow requirements which reveal the robustness of the DTC controller. We observe that when the power delivered by the PVG is not sufficient to ensure the operation of the motor, the speed of the motor is returned to zero, for this purpose, a battery will be used in a future work to stock the energy by day when solar radiation is relatively high.

6. Conclusion

In this work, we presented a model of a PV generator. Then, we developed the greenhouse climate modelling using the fuzzy logic controller to regulate temperature and relative humidity indoor greenhouse. In addition, we studied the DTC regulated ventilation motor, which is powered by a PVG, in order to decrease the overconsumption of energy. The suggested approach to energy conservation has been built on the DTC of the variable speed ventilation motor system supplied by the PVG to ensure robust control. The simulation results show that the power delivered by the PVG, shown in Figure 8, reaches a 1400 Watt when radiation is intense. The power delivered by the PVG is sufficient to ensure the operation of the motor (705 W), as shown in Figure 28. The results achieved in this paper demonstrate the effectiveness of both the fuzzy controller and the DTC control. The suggested system aims to minimize the utilisation of use of the utility of the grid.

Table 3: List of symbols, values and units

A	Windows area [50m²]
A	Constant for the renewal of the volumetric flow
Ac	Conduction and convection loss coefficient
Ai	Greenhouse area [100m²]
Cm	Heat capacity of the thermal mass
Cp	Heat capacity of the air [1003J Kg-1 °C-1]
Csat	Air saturation value
Di	Water vapor deficit in air
Ec	Evapotranspiration of crops
fogg	Water flow rate of fogging system
Fvv	Water rate in the air change flow
G	Renewal air flow
gwb	Boundary-layer conductance
gws	Stomatal conductance
hm	Coefficient of conductivity between air and thermal mass
Hi	Indoor relative humidity
Ho	Outdoor relative humidity
L	Leaves area index
MVfog	Fogging system manipulated variable

MVw	Variable for adjusting the heating system
MVα	Variable manipulated for opening windows
psat	Pressure of saturation
p	Pressure in the atmosphere [98.1 kPa]
Qcc	Conduction energy exchange and convection phenomena
Qee	Evapotranspiration of crops causes energy loss
Qff	Ground energy loss
Qmm	Exchange of energy with thermal mass
Qnn	Nebulization energy loss
Qss	Solar energy provided to the air volume
Qsm	Energy absorbed by the thermal mass by daytime
Qv	Energy exchange through window ventilation
Rn	Crop absorbs solar radiation
So	Solar radiation
Ti	Indoor temperature
Tm	Temperature of thermal mass
To	Outdoor temperature
Tref	Soil temperature at reference depth
Vi	Volume of the greenhouse [300m3]
V	Wind speed
Wh	Heating system energy
Wmax	Heating system maximum power
Xi	Absolute indoor humidity
Xo	Absolute outdoor humidity
Xsat	Absolute saturation humidity
Xsat	Absolute saturation humidity
Zref	Reference depth
α	Window opening angle
Δ	Water vapour saturation slope
γ	The psychometric constant
λ	Latent heat of vaporization
ρ	Air density [1.25 Kgair m-3]
T	Greenhouse transmission coefficient

References

[1] E. Fabrizio, 2012, Energy reduction measures in agricultural greenhouses heating: Envelope, systems and solar energy collection, Energy and Buildings, 53, 57-63.

[2] H. Challa, E.M. Nederhoff, G.P.A. Bot, and N.J. van de Brack, Greenhouse climate control in the nineties, Acta Horticulturae: 230 (1988), pp. 459-470.

[3] B.J. Bailey, and I. Seginer, Optimum control of greenhouse heating Acta Horticulturae, 245 (1989), pp. 512-518.

[4] Y. Hashimoto, Recent strategies of optimal growth regulations by the speaking plant concept Acta Hort., 260 (1989), pp. 115-122.

[5] Hahn, F.: Fuzzy controller decreases tomato cracking in greenhouses. Comput. Elect. Agric. 77, 21–27 (2011).

[6] Ferreira, P.M.; Faria, E.A.; Ruano, A.E.: Neural network models in greenhouse air temperature prediction. Neurocomputing 4, 51– 75 (2002)

[7] Rodríguez, F.; Guzmán, J.L.; Berenguel, M.; Arahal, M.R.: Adaptive hierarchical control of greenhouse crop production. Int. J. Adapt. Control Signal Process. 22, 180–197 (2008)

[8] A.J. Udink ten Cate, Modeling and (adaptive) control of greenhouse climates. Ph.D thesis. Agricultural University of Wageningen, 1983.

[9] X. Luan, P. Shi, X. Luan, P. Shi, and F. Liu, "Robust adaptive control for greenhouse climate using neural networks," Int. J. Robust Nonlin Control, vol. 21, no. 7, pp. 815–826, 2011.

[10] M. Azaza, K. Echaieb, F. Tadeo, E. Fabrizio, A. Iqbal, and A. Mami, "Fuzzy decoupling control of greenhouse climate," Arab J Sci Eng [2015] 40:2805-2812, 2015.

[11] P. Salgado and J. B. Cunha, "Greenhouse climate hierarchical fuzzy modelling," Control Eng. Pract., vol. 13, no. 5, pp. 613–628, May 2005.

[12] M. nachidi, F. Rodriguez, F. Tadeo, and J.L. Guzman, "Takagi-sugeno control of nocturnal temperature in greenhouses using air heating," ISA Transaction 50 [2011] 315-320, 2011.

[13] M. E. Ghoumari, H. J. Tantau, and J. Serrano, "Non-linear constrained MPC: Real-time implementation of greenhouse air temperature control," Comput. Electron. Agric., vol. 49, no. 3, pp. 345–356, 2005.

[14] A. L. Mohamadein, R. Hamdy and Shady M. Gadoue, "A Comparison Between Two Direct Torque Control Strategies For Flux And Torque Ripple Reduction For Induction Motors Drives", Proceedings of the Ninth International Middle East Power Systems Conference (MEPCON'2003), Shebeen Al-Koum, Egypt, December 2003, 16-18.

[15] M. G. Villalva, J. R. Gazoli, and E. R. Filho, "Comprehensive Approach to Modeling and Simulation of Photovoltaic Arrays," Power Electronics, IEEE Transactions on, vol. 24, 2009, pp. 1198-1208.

[16] M. JOMAA, M. Yassine Allani, F.TADEO, and Abdelkader Mami, Design and control of the hybrid system PV-Wind connected to the DC load, 9th International Renewable Energy Congress (IREC), 2018, 10.1109/IREC.2018.8362515.

[17] V. Salas, E. Olıas, A. Barrado, A. Lazaro, "Review of the maximum power point tracking algorithms for standalone photovoltaic systems". Solar Energy Materials & Solar Cells (2006); 90:1555–1578.

[18] H. Othmani, D. Mezghani, A. Belaid, A. Mami," New Approach of Incremental Conductance Algorithm for Maximum Power Point Tracking Based on Fuzzy Logic", International Journal of Grid and Distributed Computing Vol. 9, No. 7 (2016), pp.121-132. 2016.

[19] Hegazy Rezk, and Ali M. Eltamaly, A comprehensive comparison of different MPPT techniques for photovoltaic systems, Solar Energy, February 2015, Vol. 112:1-11.

[20] M. Jomaa, F. Tadeo and A. Mami, "Modelling and control of greenhouses powered by a renewable energy system," 2017 18th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering [STA], Monastir, 2017, pp. 482-487.

[21] X. Blasco, M. Martinez, J. M. Herrero, C. Ramos, and J. Sanchis, "Model-based predictive control of greenhouse climate for reducing energy and water consumption," Computers and Electronics in Agriculture 55, 2007, 49-70.
[22] M. Abbes, A. Farhat, A. Mami, and G. Dauphin-Tanguy, "Pseudo bond graph model of coupled heat and mass transfers in a plastic tunnel greenhouse," simulation modelling practice and theory 18, 2010, 1327-1341, 2010.
[23] Bennis, N.; Duplaix, J.; Enéa, G.; Haloua, M.; Youlal, H.: Greenhouse climate modelling and robust control. Comput. Electron.Agric. 61, 96–107 (2008)
[24] Iliev, L.; Zakeri, A.; Sazdov, P.; Baytelieva, A.M.: A fuzzy logic based approach for integrated control of protected cultivation. World Appl. Sci. J. 24, 561–569 (2013)
[25] R. C. Miranda, E. V. Ramos, R. Peniche-Vera, and G. Herrera-Ruiz, "Fuzzy Greenhouse Climate Control System based on a Field Programmable Gate Array," Biosystems Engineering, 2006, 94, 165-177.
[26] Dayan, mulation and control of ventilation rates in greenhouses. Mathematics and Computers in Simulation, 3-17, 2004.
[27] S. Z. Hassan, H. Li, T. Kamal, F. Mehmood, "Fuzzy embedded MPPT modeling and control of PV system in a hybrid power system" 12th International IEEE Conference on Emerging Technologies (ICET) 18-19-October 2016
[28] S. Z. Hassan, H. Li, T. Kamal, M. Q. Abbas, M. A. Khan and G. M. Mufti, "An Intelligent Pitch Angle Control of Wind Turbine", 2017 IEEE International Symposium on Recent Advance on Electrical Engineering (RAEE'17), Institute of Engineering and Applied Sciences, Islamabad, Pakistan from October 24-26, 2017.
[29] Villarreal, Simulated performance of a greenhouse cooling control strategy with natural ventilation and fog cooling. Biosystems Engineering, 217–228, 2012.
[30] J. Chen, Estimation of motor startup speed profile using low-resolution timing signals and motor speed-torque curve, In 20th International Compressor Engineering Conference, Purdue, Paper 1808, 2010.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 394-400 (2019

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

Transmission Line Restoration Using ERS Structure

Sushri Mukherjee[1,*], Sumana Chattaraj[1], Md. Irfan Khan[1], Dharmbir Prasad[2], Pradip Barua[1], Harish Agarwal[1]

[1]*Supreme & Co. Pvt. Ltd., 700020, West Bengal, India*

[2]*Asansol Engineering College, 713305, West Bengal, India*

ARTICLE INFO

Article history:
Received: 14 April, 2019
Accepted: 25 September, 2019
Online: 28 October, 2019

Keywords:
ERS, Bypass line
Reconductoring
Modular structure, Emergency Restoration
Power transmission services

ABSTRACT

In the past decades, various ideas have proposed for reconductoring of overhead lines to enhance power transmission capacity considering growing energy demand. Conventional method of reconductoring is use to take couple of weeks but availability of shutdown for a prolong period is a major constraint in line uprating. Emergency restoration structure (ERS) is effective to deliver such challenging task within hardly 4-5 hours line shutdown only. This quick restoration and easiness is mobility features are making ERS versatile for restoration even after any natural disasters. In this article, a case study of 132 kV restorations is described as a conceptual proof of concepts. Successful execution of this project resolved a 400 kV transmission line hindrance between Atur and Indira Puram since March 2017 and was likely to remain idle due to a 132 kV single circuit transmission line crossing its route.

Nomenclature

A_e	effective surface area of structure
B	base width of ERS
b	arm length of ERS
C_a	amplification factor
C_f	overall force co-efficient
C_F	shape co-efficient
C_g	gust response factor
E_i	expenditures
F	wind force
H	structure height
H_G	wind gradient
i	years count
k_0	factor of conversion
k_1	probability factor of risk
k_2	factor of terrain roughness
k_3	factor of site topography
m	overturning moment
p	surface contact pressure
P_s	ultimate pull out force
P_t	anchor surrounding force
P_u	shearing stress
r	discounting rate
R_i	factor of probabilistic cost
S	conductor stringing load
S_T	total structural area
V_G	wind speed
V_h	wind speed at certain height
V_H	wind speed above ground level
V_b	regional basic wind speed
V_d	design wind speed
V_r	reference wind speed
w	failure zone effective soil weight
α	wind speed inequality co-efficient
φ	structural solidity ratio
ρ	air density
$1/\lambda$	increase in velocity exponent with height

[*]Sushri Mukherjee, Research Engineer, Supreme & Co. Pvt. Ltd., W.B., India, Contact No: 8372834629 & Email ID: sushri.engg@supreme.in

ERS	emergency restoration structure
FoS	factor of safety
MDD	maximum dry density
MSPT	maximum bearing test of soil
NPV	yearly expenditures net present value
OMC	optimum moisture content of soil
RoW	right of way
R&D	research and development
SPT	standard penetration test
T&D	transmission and distribution
UTS	ultimate tensile strength

1. Introduction

1.1 General

In the era where power became basic necessity, it becomes essential for utilities to strengthen their infrastructure for accommodating the changes posed by deregulation and transfer reliable power to the consumers [1]. Pressure on the existing network to transfer more power has led to transmission line congestion. These issues could be addressed by uprating as it allows considerable increase in power transfer capability within the existing infrastructure. It's troublesome as shut down is required for reconductoring as these lines are invariably over loaded. The primary concern in emergency restoration is to revive the transmission network and get the line charged and back into operation within the stipulated time frame. These are the key objectives that compelled the development of an Emergency Restoration Structure (ERS).

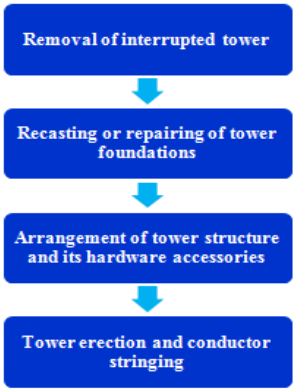


Figure 1: Flowchart of power restoration by conventional method

The Indian Power Grid has developed more complex after the resolution of "One Nation One Grid" by the Central Government. The T&D industries are enduring with R&D challenges like - reliability maintenance, useful life expansion, enhancement of transmission capacity by upgrading/uprating and public & employee safety [2-4]. The overhead transmission lines are covering miles of distance thus; it is exposed to several weather related calamities (*e.g.,* speedy wind, heavy rain, cyclone *etc*) throughout the wide area [5-7]. Unavailability of electricity due to any of these reasons or may be due to regular maintenance purpose may cause inconvenience to the consumers' along with financial loss to the service provider [8-16]. These issues could be resolved by improving grid resiliency using ERS bypass arrangement. The proposed ERS usage less time and RoW in compare conventional method of restoration considering following steps (as presented in Figure 1).

Now-a-days, most of the transmission lines are functioning inefficiently due to decay in the transmitting property. Rejuvenation of aging transmission infrastructure is required to boost its efficiency. The key challenges faced by any power transmission system are as below:

- *Transmission line congestion*
- *Fewer private players*
- *Lack in investments*
- *Strategic planning unavailability*

Transmission line utilities are always trying to reduce customer minutes lost and to avoid any concurrent financial losses [17]. They are always investing in quick restoration technology; even they are getting their manpower trained to counter devastation on emergency basis caused by natural disasters. Previously, such investment was required on a case specific critical components inventory. This led to a major chunk of financial reserves availability to handle such situations [18, 19]. The interchangeable modular ERS structure could be easily transported to even difficult locations and can be reused in future exigencies. The ERS could reduce the inventory of spare towers substantially by using it as temporary service. The installation of ERS was being optimized to overcome site challenges and reduce labour effort with rigorous installation and minutely observing the safety parameter as well enhancing the speed of the process. In this paper, a case study of 400 kV transmission line between Atur and Indira Puram, abandoned since March 2017 and was likely to remain idle due to a 132 kV Single Circuit Transmission line crossing its route.

1.2 Motivation behind the Present Study

Pressure on the existing network to transfer more power has led to transmission line congestion. These issues could be addressed by uprating as it allows considerable increase in power transfer capability within the existing infrastructure. It's troublesome as shut down is required for reconductoring as these lines are invariably over loaded. The primary concern in emergency restoration is to revive the transmission network and get the line charged and back into operation within the stipulated time frame. These are the key objectives that compelled the development of an ERS tower [20-22].

1.3 Paper Layout

In continuation of aforementioned discussions, remaining study has been presented in succeeding sections. Section 2 presents development of modular ERS structure; while Section 3 deals with reconductoring using ERS, Section 4 focused on overhead transmission line restoration; Section 5 presents observation of the Case Study and finally, the paper has been concluded in Section 6.

2. Development of ERS Module

A quick restoration of power using ERS tower could be made based on certain aspects [23-24], which are as under:

2.1 ERS Design Considerations

A case study has been done on an executed project for UPPTCL. In this paper, the guyed ERS towers were deployed for

reconductoring purpose under influence of varying wind speed in between the wind zone 30-55 m/s up to30 m height of tower as per IS:875-1987(Part III).The proposed structure sustained a major bending moment of lateral forces under influence of both tensile and compressive forces. There is optimal allocation of guy wires are required to gain stable workability of the structure. The project was successfully delivered on 5th June, 2017. The whole study could be co-related with following formulations:

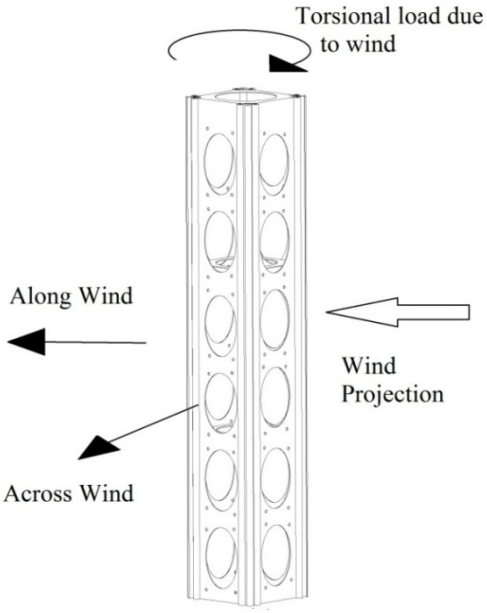


Figure 2: ERS structures' isometric view

According to metrological requirements of the site, the referred wind speed [25-28] could be given by (1),

$$V_r = \frac{V_b}{k_0} \tag{1}$$

Designed wind speed [25-31] of the ERS set up could be given in (2),

$$V_d = V_r k_1 k_2 k_3 \tag{2}$$

As per Devenport formulations for wind gradient is given by (3) [25-28]. Torsion wind load is shown in Figure 2,

$$V_H = V_G \left(\frac{H}{H_G} \right)^{\alpha} \tag{3}$$

The sectional view shows the side view of the single plane which has oval hollow section to optimize the weight and reduce effective wind pressure on the section. In order to prevent deformation of the structure several construction optimization was done. Significant features of ERS have been briefed as below:

- Erection is very quick due to modular and interchangeable section.
- Minimum numbers of component and identical fittings to reduce identification and segregation process.
- Quick installation with minimum power interruption and cost effective.
- No need of special foundation or tools.
- Erection is possible in any kind of soil and terrain.
- Compatible with 66 kV-765 kV stringing.

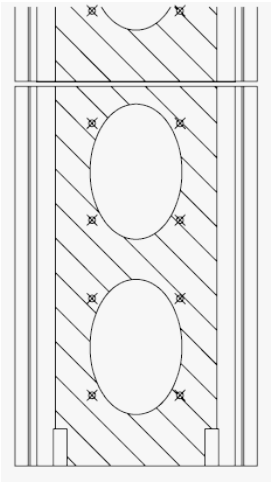


Figure 3: Sectional front view of the ERS module

Wind force [29] imposing on ERS module is given by (4),

$$F = C_f p A_e \tag{4}$$

Dynamic pressure [29], considering design variables is formulated by (5),

$$p = \frac{C_g C_F C_a \rho V_h^2}{2} \left(\frac{H}{h} \right)^{\frac{1}{\lambda}} \tag{5}$$

As per Ryle's formula, the ERS base width could be presented by equation (6) [29],

$$B = 0.013\sqrt{m} \tag{6}$$

For maintain FoS at $32^{o}\mathbf{C}$, the conductors' tension without any external load should not exceed 35% of the specified UTS. However, this percentage shrinks to 25% under loaded condition. Aspect ratio is 65.93 of the ERS modular tower as shown Figure 3 and it can be formulated as in (7) [29],

$$\varphi = \frac{S_T}{Hb} \tag{7}$$

Pull out forces to evaluate anchoring of ERS can be formulated by equation (8),

$$P_u = P_s + P_t + w \tag{8}$$

As per CIGRE, the costing NPV of the undertaken project could be formulated by (9) [30, 31],

$$NPV = \sum_{i=0}^{n} \frac{(E_i + R_i)}{(1+r)^i} \tag{9}$$

2.2 ERS towers' Hardware Integration

The ERS tower comprises of multiple components, each of these (as portrayed in Figure 4) are equally important and designed for meeting emergency requirements of the overhead

line. Each modular columns (height: 3 m) are identical and fabricated of high strength steel, so they can be easily mounted without following any particular sequence of the complete tower. Foundation plates are designed such it can allow distribution of weight to the ground with minimum load bearing capacity. There was one articulated gimbals joint, which allow linking between foundation plate and sections. The composite insulators were used to reduce RoW issues although at higher voltage levels.

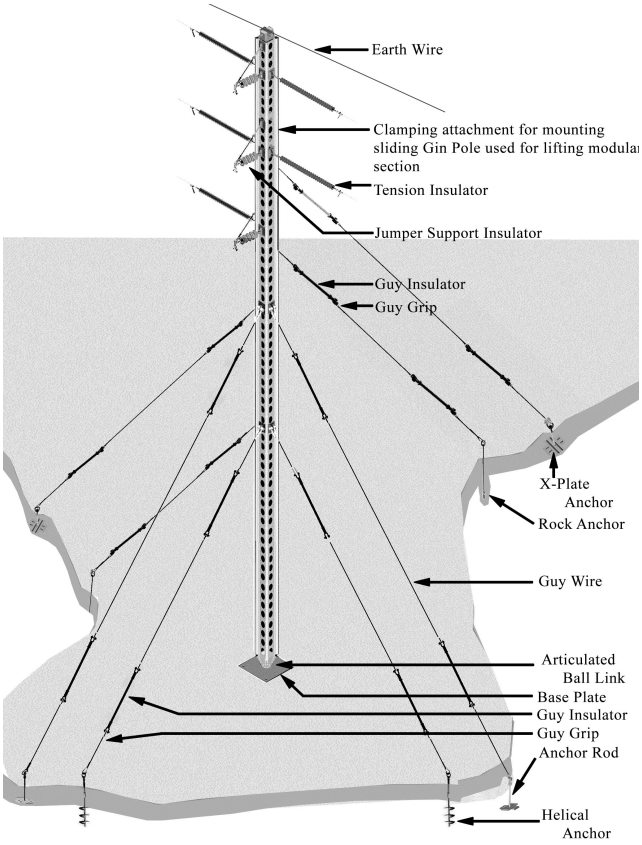


Figure 4: Generalized ERS Set Up.

3. Overhead Transmission Line Restoration

This paper briefly presents a case of electricity transmission from Loni to Mohan Nagar division (UP, India) under following subsections:

3.1 GIS Mapping of the Site

The ERS tower has been utilized at UPPTCL site to re-build 132 kV lines confronting flyover, water passage and street intersections, separately (depicted in Figure 5). However, 132 kV overhead lines was thwarting hanging of 400 kV transmission lines, as height of that line was 51 m from the top most earth wire.

Consequently, a restructuring was made to curtail height of 132 kV by de-amassing second tower (ref to Figure 5) such that, the 400 kV line can be set back on it. Total three ERS towers have been utilized to bypass. Usually, three weeks supply interruption could have been there with conventional method for restoration but application of ERS has delivered this project by 5 hours only. Before initiating execution work of this project, a extensive site survey and planning was done to accomplish this project.

Figure 5: Ghaziabad site restoration layout

3.2 Deployment of Restoration Works

In the current work, three ERSs (height: 30m) were deployed based on the requirements of the specific site UP state, in order to meet resilience and sustainability of bypass arrangement [28]. The analysis of stringing arrangement was performed using PLS CADD software tools. The complete solutions for line restoration or construction of any kind of bypass line by using ERS tower on rental or end to end services basis is being provided, as we have a dedicated team of execute the work. It involves several steps for the enactment of the ERS installation:

Step-1: The initial step involves site survey of the installation site. There are various factors that are taken into account before the installation. A dedicated team is at the first analyses and collects information regarding site. It includes the information such as accessibility to the site, existing tower to ERS tower clearance, RoW clearance, layout of the site, span length *etc*. Site accessibility also helps to pre-determine the anchor to be used for the type of soil as well kind of tools. After the examination process and collection of all the details, the project engineers proceed towards the next step. The load assessment is done by the design team as per the side condition and feasibility.

Step-2: The second step involves material transportation such as the modular sections and all the equipment that are required for the proper guying and installation of the tower.

Step-3: The third step includes the installation part. Most importantly the major advantage of our ERS towers is that civil foundation is not required. The tower is entirely supported using guy wires. Hence, it is essential that all the guy wires are properly anchored to the soil. Our team took this as a challenge and developed different kind of anchors that would support the structure for different soil conditions.

Step-4: After the installation of ERS base the gimbals joint is then attached to it and with the help of it modular section is installed. Gin poles are used for lifting the next modular sections one by one.

Step-5: For stabilizing the tower guy arrangement has been used. After that all the hardware are installed and conductor stringing is done. Some extra wires are removed to ease the congestion along the axis of the conductor since the

tensioned conductor support the tower followed by inspection and joint commissioning to charge the line.

The overhead transmission line has been circumventing from tower-1 to that of -2 by a new root as mapped in Figure 6(a) and erected line was shown in Figure 6(b).

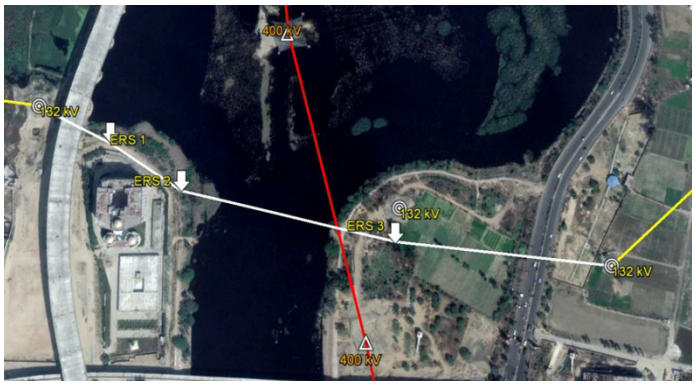


Figure 6(a): Bypass arrangement routing layout

Figure 6(b): Flyover and river crossing

The range existences between re-steered lines are customary cross section tower-1 to ERS-1: 111 m, ERS-1 to -2: 93 m, ERS-2 to -3: 229 m lastly ERS-3 to regular lattice tower-3: 250 m. In continuation with site overview, skilled team with their tools deployed to pursue the project.

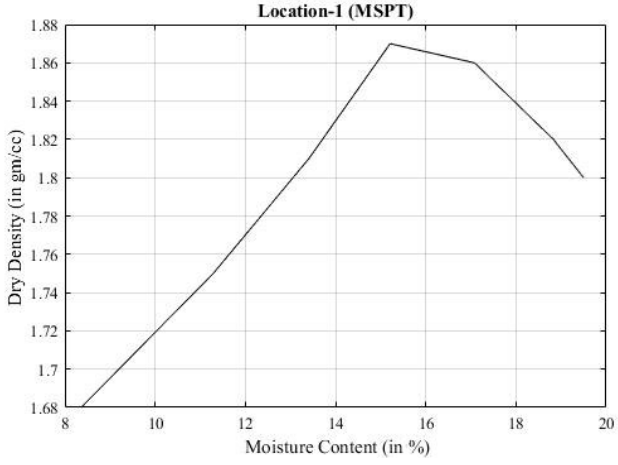


Figure 7: Characteristics of MSPT test for soil sample-1.

4. Observations from Case Study

The heavy compaction test has been tested for soil samples taken from location-1 and -2 while; light compaction test is found to be appropriate for location-3. These types of test were carried out to check MDD and OMC of soil as per IS: 2720 (Part VIII) 1983. The field dry density core cutter testing method complying IS:2720 (Part VII & XXIX) has been performed across the field at each location as shown in Table 1.

Table 1: Field test findings from core cutter method.

Parameters	Sample- 1	Sample- 2	Sample-3
Field BD (gm/cc)	1.93	1.91	1.92
Field MC (%)	4.79	6.16	4.82
Field DD (gm/cc)	1.84	1.80	1.83
DD (% of MDD)	98.39%	96.25%	97.86%

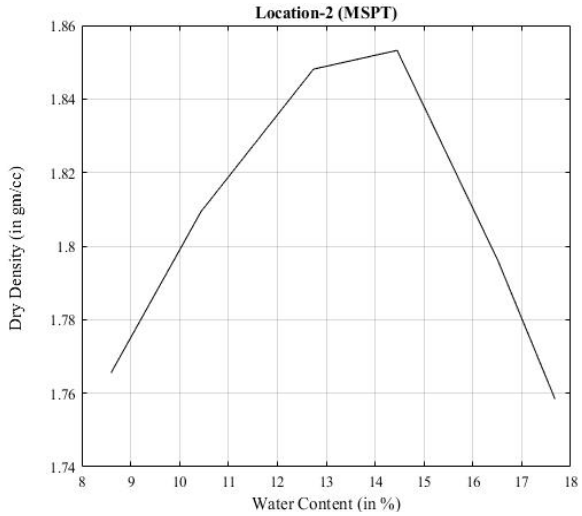


Figure 7: MSPT test characteristics of soil sample-2.

The recorded test findings of moisture content and dry density are provided in Table 2 and their graphs are plotted in Figure 7-9. The interesting outcomes (OMC & MDD) are highlighted by **boldfaced** in the subsequent column of Table 1. The compaction test observations exceeding 95% of the MDD; thus, the site was viable to erect the ERS tower.

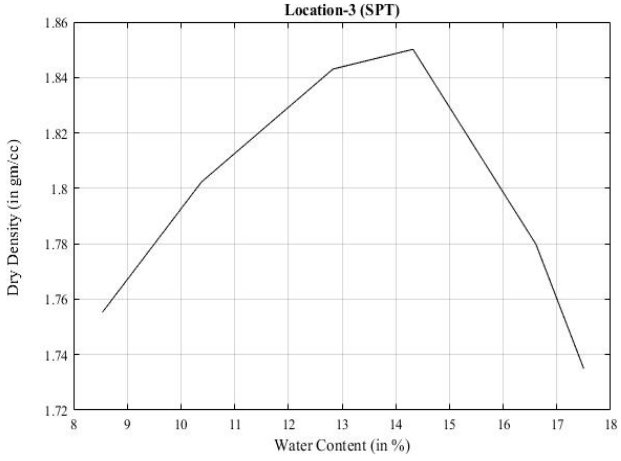


Figure 8: SPT test characteristics of soil sample-3

Table 2: Results obtained after compaction test

Locations	Testing methods	Parameters	
		Content of moisture (%)	Dry density (gm/cc)
1	MSPT	8.360	1.68
		11.29	1.75
		13.41	1.81
		15.21	**1.87**
		17.08	1.86
		18.83	1.82
		19.50	1.80
2	MSPT	8.600	1.766
		10.44	1.809
		12.74	1.848
		14.45	**1.853**
		16.51	1.796
		17.68	1.758
3	SPT	8.540	1.76
		10.38	1.80
		12.83	1.84
		14.32	**1.85**
		16.61	1.78
		17.50	1.73

Table 3: Site specific benefit comparison

S. No.	Particulars	Restoration using ERS	Conventional restoration
1	Structural footprint	<1 m^2	36-64 m^2
2	Foundations of structure	Not required.	Depth: 2-4 m
3	Angle of protection	65º	30º
4	Down time of supply	5 Hours	20 Days
5	Average value of power flow	5 MW	5 MW
6	Generation loss/day	120000 Units	120000 Units
7	Unit price	INR 2.5	INR 2.5

After execution of this project, a cost-benefit analysis was made (given in Table 3) to ascertain future scope of ERS in service customers.

5. Conclusion

Indeed, even in open vitality markets, electricity transmissions are normally considered as a usual monopoly and sometimes bring contests among investors. This starts leapfrogging of new ideas so as to get by in the aggressive power industry. Thusly, development of ERS gives an elective potential answer for detour the customary transmission tower for rebuilding just as line reconductoring. Line uprating proves to be a smart solution to cater the increased loads on existing transmission lines. However, this exercise is problematic as the shutdown required for reconductoring cannot be provided as those lines are already critical. This exercise helped in completing a project which would evacuate 580 MW lines where substantial investment was already made on 98% stretch of the line and associated substation would have remained idle for indefinite period. Project milestones and achievements are concluded as below:

- *Line Crossing:* Stringing was challenging as the line had a flyover, a river and a highway crossing
- *Delay in Shutdown:* Even after installing the ERS, re-tapping was frenetic due to shutdown required for 2 hours from UPPTCL was difficult.
- *No Heavy Machinery:* No heavy machinery or earth movers were introduced for execution of above, other than for transportation, which depicts the agility of the system.
- *Challenging Feat:* The bypass line across the highway was done without any disruption of highway traffic even for a minute.
- *Rough Weather:* There was stormy wind after erection but no damage was made to the ERS.

Conflict of Interest

Authors are declaring that there is no conflict of interest with any other individual or organization.

Acknowledgement

The project shown as a case study was done for UPPTCL, India. Authors are thankful to Mr. Arup Das (SCPL) and Mr. Prakash Reddy (SCPL) for their active co-operation.

References

[1] "World Energy Outlook Special Report : India Energy Outlook" International Energy Agency, 2015.

[2] O. H. Abdalla, T. M. Alkhusaibi, M.A. Thani, M. N. Alkhusaibi, "Restoration of a 132 kV Over Head Transmisssion Line affected by Tropical Cyclone Gonu in Oman" in IEEE Proc. Transm. Distrib. Conf. and Expo., 1-5, 2008. DOI: 10.1109/TDC.2008.4517063

[3] L. N. Agrawal, P. E. Erickson, "Planning and Training reduce Restoration time for damaged transmission lines in Inida" in Int. Proc. on Transm. and Distrib. Constr., Oper.and Live-Line Maint., 101-108, 2000. DOI: 10.1109/TDCLLM.2000.882807

[4] M.I. Khan, D. Prasad, K. Sarkar, S. Mukherjee, P. Barua, H. Agarwal, 2019, March. Finite Element Modeling of 400 kV ERS Bypass Transmission Line Quantifying Installation Difficulty. In 2019 IEEE PES GTD Grand International Conference and Exposition Asia (GTD Asia) (pp. 526-531). IEEE. 10.1109/GTDAsia.2019.8715871

[5] D. Prasad, R. Kumaran, G. Agarwal, P.K. Reddy, Md. I. Khan, 2016. Emerging Solutions for Line Compaction. CIGRE India Journal, 5(2), pp. 21-26.

[6] S. Mukherjee, D. Prasad, Md. I. Khan, P. Barua, H. Agarwal, 2019. Hexa Spacer Damper for Vibration Energy Decaying of 765 kV Transmission Line. 2nd International Conference on Innovations in Power and Advanced Computing Technologies, i-PACT-2019, on 22nd - 23rd March, 2019 at Vellore Institute of Technology, Vellore, India.

[7] D. Prasad, V. Mukherjee, 2016. A novel symbiotic organisms search algorithm for optimal power flow of power system with FACTS devices. Engineering Science and Technology, an International Journal, 19(1), pp.79-89. 10.1016/j.jestch.2015.06.005

[8] D. Prasad, A. Mukherjee, V. Mukherjee, 2017. Application of chaotic krill herd algorithm for optimal power flow with direct current link placement problem. Chaos, Solitons & Fractals, 103, pp. 90-100. 10.1016/j.chaos.2017.05.037

[9] D. Prasad, A. Mukherjee, G. Shankar, V. Mukherjee, 2017. Application of chaotic whale optimisation algorithm for transient stability constrained optimal power flow. IET Science, Measurement & Technology, 11(8), pp.1002-1013. 10.1049/iet-smt.2017.0015 10.1049/iet-smt.2017.0015

[10] D. Prasad, A. Mukherjee, V. Mukherjee, 2017. Transient stability constrained optimal power flow using chaotic whale optimization algorithm. In Handbook of neural computation (pp. 311-332). Academic Press. 10.1016/B978-0-12-811318-9.00017-X
[11] D. Prasad, V. Mukherjee, 2018. Solution of optimal reactive power dispatch by symbiotic organism search algorithm incorporating FACTS devices. IETE Journal of Research, 64(1), pp.149-160. 10.1080/03772063.2017.1334600
[12] V. Mukherjee, A. Mukherjee and D. Prasad, 2018. Whale Optimization Algorithm with Wavelet Mutation for the Solution of Optimal Power Flow Problem. In Handbook of Research on Predictive Modeling and Optimization Methods in Science and Engineering (pp. 500-553). IGI Global. 10.4018/978-1-5225-4766-2.ch023
[13] D. Prasad, A. Banerjee, R.P. Singh, 2019. Optimal Reactive Power Dispatch Using Modified Differential Evolution Algorithm. In Advances in Computer, Communication and Control (pp. 275-283). Springer, Singapore. 10.1007/978-981-13-3122-0_26
[14] A. Banerjee, D. Prasad, R.P. Singh, 2015. Smart reactive power control of hybrid power system model by Seeker Optimized Algorithm. In ETES-2015 conference organized by IET on 16-17th July, 2015 at Asansol, West Bengal, India.
[15] R.P. Singh, A. Banerjee, D. Prasad, 2014. Solution of optimal power flow problem using modified differential evolution algorithm. In ETES-2014 conference organized by IET at Asansol, West Bengal, India.
[16] R.P. Singh, D. Prasad, P. Kumar, and M.I. Khan, 2017, August. Modified differential evolution algorithm for optimal power flow with FACTS devices. In 2017 International Conference on Smart grids, Power and Advanced Control Engineering (ICSPACE) (pp. 48-52). IEEE. 10.1109/ICSPACE.2017.8343404
[17] N. Ozog, E. Desjardins, J. Jatskevich, "Bulk power system restoration interdependency risk modelling" in IEEE Proc. Elect. Power Energy Conf., 1-7, 2008. DOI: 10.1109/EPC.2008.4763347
[18] R. J. Schweiner, K. E. Twomey, K. E. Lindsey, "Transmission Line Emergency Restoration Philosophy at Los Angeles Department of Water and Power," in IEEE Proc. Transm. and Distrib. Constr., Oper. Live-Line Maint., 11-17, 2003. DOI: 10.1109/TDCLLM.2003.1196462
[19] J. C. Pohlman, "The case for Universal Standards to cover the structural components and practices used in Emergency Restoration Systems" in Int. Proc. on Transm. and Distrib. Constr., Oper.and Live-Line Maint., 109-115, 2000. DOI: 10.1109/TDCLLM.2000.882808
[20] J. F. Nolasco, P. Nefzger, U. Kaintzyk, Overhead power lines: Planning, Design, Construction, Springer, 2003.
[21] T. S. Kishore, S. K. Singal, "Optimal economic planning of power transmission lines: A review" Ren. and Sust. Energy Rev., 39(5), 949-974, 2014.https://doi.org/10.1016/j.rser.2014.07.125
[22] L. Bauer, P. Ulardic, J. Muller, "Reinforcement Strategies for Extending the Service Life of 110 kV Overhead Transmission Lines" in Int. Conf.and Exhibition on Electr. Distr., 1997. DOI: 10.1049/cp:19970547
[23] S. G. Krishnasamy, G. L. Ford, C. I. Orde, "Predicting the structural performance of transmission lines uprated by reconductoring" IEEE Trans. Power Appar. Syst., 100(5), 2271-2277, 1981. DOI: 10.1109/TPAS.1981.316740
[24] "Conductors for the Uprating of Existing Overhead Lines,"CIGRÉ WG B2.55 Brochure (Ref. No. TB 244), 2003.
[25] A. Marin, J. Salmeron, "Electric capacity expansion under uncertain demand:decomposition approaches" IEEE Trans. Power Syst., 13(2), 333-339, 1998. DOI: 10.1109/59.667347
[26] I. Albizu, A. J. Mazón, I. Zamora, "Methods for Increasing the Rating of Overhead Lines" in Power Tech IEEE Proc., Russia, 1-6, 2005. DOI: 10.1109/PTC.2005.4524481
[27] M. T. Chay, F. Albermani and H. Hawes, "Wind loads on transmission line structures in simulated downbursts," in First World Congress on Asset Management, Gold Coast, Australia, 2006.
[28] S. S. Murthy, A. R. Santhakumar "Transmission line structures," McGRAW-HILL Book Co., 1990.
[29] P. J. Ryle, "Steel tower economics" J. Inst. Electr. Eng. - Part II: Power Eng., 93(33), 263-274, 1946. DOI: 10.1049/ji-2.1946.0063
[30] H. Niroumand, Soil reinforecement for anchor plates and uplift response, Butterworth-Heinemann, 2017.
[31] J. C. Kaltenbach, J. Peschon, E. H. Gehrig "A mathematical optimization technique for the expansion of electric power transmission systems" IEEE Trans. Power Appar. Syst., 89(1), 113-119, 1970. DOI: 10.1109/TPAS.1970.292677

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 401-408 (2019)

www.astesj.com

Special Issue on Advancement in Engineering and Computer Science

ASTESJ

ISSN: 2415-6698

EKMC: Ensemble of kNN using MetaCost for Efficient Anomaly Detection

Niranjan A[*,1], Akshobhya K M[1], P Deepa Shenoy[1], Venugopal K R[2]

[1]*Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore University, Bangalore, India.*

[2]*Bangalore University, Bangalore, India.*

ARTICLEINFO

Article history:
Received: 31 May, 2019
Accepted: 29 September, 2019
Online: 28 October, 2019

Keywords:
Anomaly Detection
kNN
MetaCost

ABSTRACT

Anomaly detection aims at identification of suspicious items, observations or events by differing from most of the data. Intrusion Detection, Fault Detection, and Fraud Detection are some of the various applications of Anomaly Detection. The Machine learning classifier algorithms used in these applications would greatly affect the overall efficiency. This work is an extension of our previous work ERCRTV: Ensemble of Random Committee and Random Tree for Efficient Anomaly Classification using Voting. In the current work, we propose SDMR a simple Feature Selection Technique to select significant features from the data set. Furthermore, to reduce the dimensionality, we use PCA in the pre-processing stage. The EKMC (Ensemble of kNN using MetaCost) with ten-fold cross validation is then applied on the pre-processed data. The performance of EKMC is evaluated on UNSW_NB15 and NSL KDD data sets. The results of EKMC indicate better detection rate and prediction accuracy with a lesser error rate than other existing methods.

1. Introduction

A process involving the identification of data points that do not fit with the remaining data points is referred to as Anomaly Detection. Hence, Anomaly Detection is employed by various machine learning applications involving the Detection of Intrusions, or Faults, and Frauds. Anomaly Detection can be achieved based either on nature of data or circumstances. There are three approaches for Anomaly Detection used under different circumstances: Static Rules approach, when the Training data is missing and when the Training data is available.

1.1. Static Rules Approach

In this approach, a list of known anomalies is identified, and rules are written to identify these anomalies. Rules are generally written using pattern mining techniques. Since identification of Static Rules is complex, machine learning approach that involves automatic learning of the rules is preferred.

1.2. When Training Data is missing

When the data set lacks a class label, we may use Unsupervised or Semi supervised learning techniques for Anomaly Detection. However, evaluating the performance of this approach is not possible because there shall be no test data either.

1.3. When Training Data is available

Even while the training set is available, the number of Anomaly samples will be too less when compared to the benign samples and hence there shall be class imbalance in such data sets. To overcome this problem, new sets are created by resampling data several times.

Anomaly detection can happen only after a successful classification. The efficiency of Anomaly Detection applications therefore depends on the classifiers used. Prediction Accuracy, ROC Area and Build time are some of such metrics that can measure the efficiency of a classifier. They are in turn based on Detection Rate (DR) and False Positive Rate (FPR). While DR is the correctness measure, FPR is the incorrectness measure during classification. ROC(Receiver Operating Characteristic) is a graphical representation of the ability of a binary classifier system obtained by varying its threshold. ROC involves plotting of TPR values (Y-axis) against FPR values (X-axis) at different threshold values. The time taken to train the given model is its build time.

[*]Niranjan A, +91 8123033210, a.niranjansharma@gmail.com

Any work is expected to have maximum value for DR and least or nil values for FPR, Error rates and Build time. This work focuses on the selection of features that are significant, from the data sets and reduction in their dimensionality while maintaining the detection accuracy. To achieve this, classifier algorithms with better individual performances are determined and are experimented with various combinations (ensemble) of classifiers. It was observed through our experiments that kNN offers best results in terms of the chosen metrics.

kNN is a typical classifier that learns based on instances. It is often referred to as a Lazy learning algorithm, because it defers computation until actual classification. The kNN algorithm assumes that similar things exist in proximity and therefore a sample from the test set is classified based on the predictions made by most of its neighbors.

Bagging, Boosting, Voting and Stacking are the ensembling techniques available today. The Bagging approach draws *n* instances randomly from a training set using a distribution that is uniform and learns them. The process is repeated several times. Every repetition generates one classifier. Boosting, a similar approach as that of bagging, focuses more on instances that were learnt incorrectly and monitors the performance of the machine learning algorithm. After constructing several classifiers in this manner, it performs a vote of the weights associated with the individual classifiers for making the final prediction. Each classifier is assigned weights based on its achieved detection accuracy on its training set. Voting requires the creation of several sub-models, allowing each of them to vote on the outcome of prediction. Stacking involves the training of different learning algorithms on the available data and providing the predictions of each learning algorithm as additional inputs to the combiner algorithm for the final training. In StackingC, Linear Regression is used as the Meta Classifier. A way of representing a linear equation by merging a set of input values (x) that are numeric into a predicted output value (y), may be defined as Linear Regression. This work involves an ensembling technique for the classification of the test samples present in the data set using MetaCost. MetaCost would produce results that are like the one that is created by passing the base learner (kNN in our case) to Bagging, which eventually is passed to a Cost Sensitive Classifier that operates on least expected cost. The only difference that we can observe is that MetaCost generates only one cost-sensitive classifier of the base learner, offering fast classification and interpretable output. This implementation uses all iterations of Bagging by reclassifying the training data.

Our experiments on the two benchmark data sets namely *NSL-KDD* and *UNSW_NB15*, prove that an ensemble of kNN using MetaCost yields better results compared to various machine learning algorithms. The NSL-KDD data set comprises of 41 features, and a class label to indicate an instance as normal or anomalous. The *UNSW_NB15* data set on the other hand has 44 features plus one class label.

In this extension work [1], we propose SDMR for Feature Selection that exploits the advantages of various existing Weight Based Ranking Algorithms. In addition to SDMR, the data set is also subjected to PCA for dimensionality reduction during the preprocessing stage. The Principal Component Analysis (PCA) when applied on a data set having many variables (features) correlated with one another, reduces its dimensionality by only retaining the variation present in it. The existing variables of the data set are transformed to a new set of variables, known as the principal components (or PCs) that are orthogonal such that the correlation between any pair of variables is 0. The resultant set is then subjected to EKMC (Ensemble of kNN using MetaCost) with a cross validation of ten-folds before recording the performance metrics.

The details of our proposed framework are provided in Sections 3 and 4, respectively.

The key contributions of this extended paper are as follows.

1. *SDMR* (Standard Deviation of Mean of Ranks) to discard all those features whose ranks are less than the computed value,

2. Use of PCA for the further reduction of dimensionality of the data set.

3. *EKMC* Framework for efficient Anomaly classification.

The remainder of this article is organized as follows: Background and previous work related to ADS and our novel EKMC technique are explained in Section 3. Section 3 also discusses about the details of the novel SDMR Feature Selection technique. Section 4 presents the experimental results and analysis of the proposed EKMC using the two benchmark data sets. Finally, we conclude our work and suggest directions for further research.

2. Background and Related Works

ERCRTV [1] that forms the base work for the current work, uses Correlation based Feature Selection (CFS) algorithm for Feature Selection from the NSL KDD and KDD CUP 99 data sets. It selects only eight prominent features from them. The data subset with only chosen features is provided to an ensembled model of Random Committee and Random Forest using Voting. A ten-fold cross validation is performed on the model before recording the performance metrics. CFS being one of the Filter based Feature Selection algorithms, is faster, but is less accurate. Hence our current work involves a simple and more efficient SDMR technique for Feature Selection and Metacost classifier with kNN as the base classifier for the classification of Anomalous and benign samples. The MetaCost classifier relabels the class feature of the training set using meta learning technique. The modified training set is then used to produce the final model.

The authors of [2], propose a novel approach involving Two-layer dimensionality reduction followed by a Two-Tier classification for efficient detection of intrusions in IoT Backbone Networks. Their approach addresses the limitations of making wrong decisions and increased computational complexity of the classifier due to higher dimensionality. Component Analysis and

Linear Discriminate Analysis form the Two Layers of Dimensionality Reduction during the preprocessing stage while Naïve Bayes and Certainty Factor variation of the K-Nearest Neighbor techniques form the Two Tiers of classification. A Detection Accuracy of 84.82 on twenty percent of the NSL-KDD training set is achieved by their work.

The methodology presented in [3] illustrates a detection technique based on anomaly detection involving data mining techniques. The paper discusses about the possible use of Apache Hadoop for parallel processing of extremely huge data sets. Dynamic Rule Creation technique that is adopted by their authors ensures that even new types of security breaches are detected automatically. The error rates of below ten percent can be observed from their findings.

The authors in their work [4], present a PSO-based feature selection followed by a two-tier ensembling model involving Boosting and Random Subspace Model (RSM). They illustrate with their results that accuracy and false positive rate (FPR) are better compared to all other models.

The work presented in [5] illustrates the importance of outlier detection in the training set that is achieved through Robust Regression technique during the preprocessing stage. Their work further proves that their model is far more superior to the normal Linear Regression technique that is used by most researchers. With their experimental data, the authors compare their model with Linear Regression Model and demonstrate that their Model is much superior especially in environments with bursty network traffic and pervasive network attacks.

The authors of [6] outline a Proactive Anomaly Detection Ensemble (ADE) technique for the timely anticipation of anomaly patterns in a given data set. Weighted Anomaly window is used as the ground truth to train the model allowing it to discover an anomaly well before its occurrence. They explore various strategies for the generation of ground truth windows. With their results, they establish that ADE exhibits at least ten percent improvement in earliest detection score as compared with other individual techniques across all the data sets that are considered for experimentation.

3. EKMC Technique

The current work revolves around Preprocessing and Classification phases. Feature Selection forms the main layer of preprocessing, since not all attributes in the data set are relevant during the analysis. We propose a novel SDMR for Feature Selection that exploits the advantages of various existing Weight Based Ranking Algorithms. In addition to SDMR, the data set is also subjected to PCA for dimensionality reduction during this phase. In the classification phase, we subject the resultant subset to the proposed EKMC algorithm with ten-fold cross validation for measuring the performance metrics. The framework of our proposed technique is depicted in Fig.1. The experiments are carried out on two benchmark data sets namely UNSW-NB15 and NSL-KDD. The NSL-KDD comprises of 125973 samples in the training and 22544 in the test set. EKMC model was trained making use of the training set and was then tested with the test set. The performance of the proposed model was further validated by running the model on UNSW-NB15 data set comprising of 175,341 records in the training set and 82,332 records in the test set.

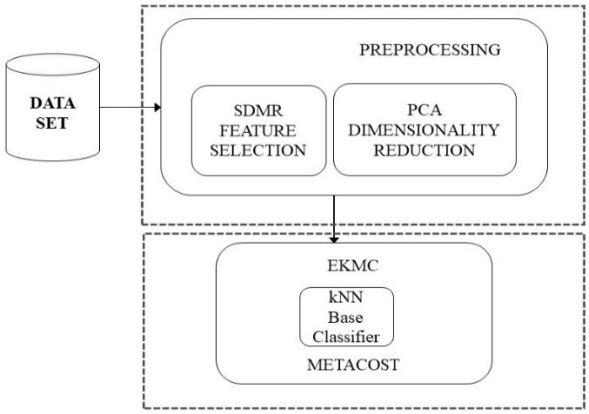


Figure 1: Proposed framework for Anomaly Detection

The various Weight based Feature Selection Algorithms that were employed to compute the Ranks R_i in the proposed SDM_R are Information Gain, Information Gain Ratio, Weight by Correlation, Weight by Chi Squared Statistics, Gini Index, Weight by Tree importance, and Weight by Uncertainty. The SDM_R that we obtained for the NSL KDD Data set was 0.278831 and that of UNSW-NB15 was 0.184325. All those features that are less than the SDM_R values were discarded from the data sets. The proposed SDMR returned only 15 features out of 41 in case of NSL KDD and 11 features out of 44 features in case of UNSW-NB15. These subsets of features of both the data sets are further subjected to PCA for dimensionality reduction. The resulting feature subsets are finally subjected to the proposed EKMC framework. The proposed EKMC algorithm for efficient Anomaly Detection is presented in Algorithm 2. Table 1 and Table 2 list the Ranks determined using different Rank-Based Feature Selection Algorithms on NSL KDD and UNSW-NB 15 respectively. The proposed SDMR[7] for Feature Selection is presented in Algorithm1. The experimental results as indicated in Table 4 suggest that the kNN classifier offers best prediction accuracy, precision, recall, F-1 measure and Detection Accuracy with least classification error value out of the 20 classifier algorithms.

Algorithm 1: SDMR (Standard Deviation of Mean of Ranks) Feature Selection

Input: D data set having n number of Features
Output: subset F of D with Most Significant Features

1. for each feature $f_i \in D$ do
   Determine Ranks R_i using different Weight Based Feature Selection Techniques
   next
2. for each feature $f_i \in D$ do
   Compute Sum ($\sum R_i$) and Mean mR_i of Ranks
   $\sum R_i = R_1 + R_2 + \ldots + R_n$ and $mR_i = \frac{\sum Ri}{n}$
   next
3. Compute Standard Deviation of Mean of Ranks SDM_R
4. Discard all $f_i \in D < SDM_R$

5. **return** *F*

Table 1: Feature Ranks of NSL KDD

Attribute	Information Gain	Information gain ratio	Correlation	Chi Squared statistics	Gini Index	Tree importance	Uncertainty
f1_duration	0.03	0.18	0.06	0.01	0.03	0.00	0.02
f2_protocol_type	0.09	0.12	0.07	0.11	0.11	0.00	0.16
f3_service	1.00	0.29	0.43	1.00	1.00	0.00	0.67
f4_flag	0.77	0.56	0.64	0.81	0.81	0.00	1.00
f5_src_bytes	0.90	1.00	0.01	0.00	0.96	1.00	0.00
f6_dst_bytes	0.85	0.94	0.00	0.00	0.89	0.94	0.00
f7_land	0.00	0.02	0.01	0.00	0.00	0.00	0.00
f8_wrong_fragment	0.01	0.22	0.13	0.01	0.01	0.00	0.04
f9_urgent	0.00	0.08	0.00	0.00	0.00	0.00	0.00
f10_hot	0.00	0.14	0.02	0.00	0.00	0.00	0.00
f11_num_failed_logins	0.00	0.10	0.00	0.00	0.00	0.00	0.00
f12_logged_in	0.60	0.69	0.92	0.64	0.64	0.00	1.00
f13_num_compromised	0.00	0.12	0.01	0.00	0.00	0.00	0.00
f14_root_shell	0.00	0.04	0.03	0.00	0.00	0.00	0.00
f15_su_attempted	0.00	0.12	0.03	0.00	0.00	0.00	0.00
f16_num_root	0.01	0.13	0.01	0.00	0.01	0.00	0.00
f17_num_file_creations	0.00	0.11	0.03	0.00	0.00	0.00	0.00
f18_num_shells	0.00	0.04	0.01	0.00	0.00	0.00	0.00
f19_num_access_files	0.00	0.12	0.05	0.00	0.00	0.00	0.01
f20_num_outbound_cmds	0.00	0.00	0.00	0.00	0.00	0.00	0.00
f21_is_host_login	0.00	0.08	0.00	0.00	0.00	0.00	0.00
f22_is_guest_login	0.00	0.03	0.05	0.00	0.00	0.00	0.01
f23_count	0.49	0.56	0.77	0.58	0.57	0.00	0.56
f24_srv_count	0.03	0.22	0.00	0.06	0.03	0.10	0.12
f25_serror_rate	0.55	0.72	0.87	0.58	0.58	0.00	0.91
f26_srv_serror_rate	0.56	0.72	0.86	0.58	0.58	0.68	0.93
f27_rerror_rate	0.08	0.15	0.34	0.09	0.09	0.00	0.15
f28_srv_rerror_rate	0.07	0.16	0.34	0.09	0.09	0.00	0.15
f29_same_srv_rate	0.70	0.82	1.00	0.77	0.75	0.79	0.94
f30_diff_srv_rate	0.67	0.77	0.27	0.03	0.74	0.73	0.06
f31_srv_diff_host_rate	0.14	0.21	0.16	0.12	0.16	0.00	0.19
f32_dst_host_count	0.22	0.26	0.50	0.27	0.27	0.00	0.25
f33_dst_host_srv_count	0.59	0.66	0.96	0.74	0.67	0.60	0.71
f34_dst_host_same_srv_rate	0.59	0.66	0.92	0.71	0.67	0.00	0.69
f35_dst_host_diff_srv_rate	0.51	0.57	0.32	0.71	0.59	0.49	0.06
f36_dst_host_same_src_port_rate	0.14	0.16	0.12	0.06	0.17	0.00	0.07
f37_dst_host_srv_diff_host_rate	0.22	0.28	0.08	0.04	0.26	0.00	0.09
f38_dst_host_serror_rate	0.56	0.74	0.87	0.59	0.57	0.70	0.91
f39_dst_host_srv_serror_rate	0.58	0.76	0.87	0.58	0.58	0.00	0.97
f40_dst_host_rerror_rate	0.07	0.14	0.34	0.09	0.09	0.00	0.13
f41_dst_host_srv_rerror_rate	0.11	0.25	0.34	0.11	0.12	0.00	0.19

Table 2: Feature Ranks of UNSW-NB 15

Attribute	Information Gain	Information gain ratio	Correlation	Chi Squared statistics	Gini index	Tree importance	Uncertainty
f1_id	0.32	0.51	0.49	0.62	0.38	0.83	0.46
f2_dur	0.20	0.29	0.04	0.00	0.24	0.13	0.01
f3_proto	0.19	0.08	0.33	0.20	0.20	0.00	0.20
f4_service	0.08	0.00	0.28	0.09	0.09	0.00	0.10
f5_state	0.27	0.27	0.65	0.32	0.32	0.30	0.39
f6_spkts	0.11	0.16	0.05	0.00	0.14	0.30	0.00
f7_dpkts	0.20	0.30	0.12	0.00	0.25	0.07	0.00
f8_sbytes	0.11	0.30	0.02	0.00	0.14	1.00	0.00
f9_dbytes	0.20	0.30	0.08	0.00	0.25	0.03	0.00
f10_rate	0.24	0.37	0.43	0.25	0.27	0.24	0.29
f11_sttl	0.46	1.00	0.80	0.54	0.53	0.90	0.70
f12_dttl	0.20	0.29	0.02	0.52	0.24	0.48	0.63
f13_sload	0.24	0.38	0.21	0.00	0.28	0.15	0.00
f14_dload	0.26	0.58	0.45	0.13	0.33	0.25	0.26
f15_sloss	0.10	0.14	0.00	0.00	0.13	0.19	0.00
f16_dloss	0.10	0.21	0.10	0.00	0.14	0.08	0.00
f17_sinpkt	0.13	0.30	0.20	0.02	0.17	0.08	0.07
f18_dinpkt	0.20	0.30	0.04	0.00	0.24	0.02	0.00
f19_sjit	0.10	0.11	0.02	0.00	0.13	0.11	0.00
f20_djit	0.10	0.10	0.06	0.00	0.13	0.10	0.00
f21_swin	0.10	0.11	0.47	0.13	0.13	0.00	0.18
f22_stcpb	0.09	0.09	0.34	0.10	0.12	0.00	0.08
f23_dtcpb	0.09	0.09	0.34	0.09	0.12	0.00	0.07
f24_dwin	0.09	0.09	0.43	0.12	0.12	0.00	0.16
f25_tcprtt	0.09	0.17	0.03	0.01	0.11	0.06	0.02
f26_synack	0.09	0.16	0.05	0.00	0.11	0.21	0.01
f27_ackdat	0.09	0.16	0.00	0.00	0.11	0.13	0.01
f28_smean	0.05	0.05	0.04	0.02	0.06	0.29	0.02
f29_dmean	0.20	0.30	0.38	0.10	0.25	0.09	0.12
f30_trans_depth	0.00	0.00	0.00	0.00	0.00	0.12	0.00
f31_response_body_len	0.01	0.10	0.02	0.00	0.01	0.00	0.00
f32_ct_srv_src	0.08	0.13	0.31	0.09	0.09	0.14	0.11
f33_ct_state_ttl	0.41	0.90	0.61	0.56	0.48	0.40	0.62
f34_ct_dst_ltm	0.09	0.17	0.31	0.09	0.09	0.09	0.13
f35_ct_src_dport_ltm	0.11	0.22	0.41	0.12	0.12	0.02	0.19
f36_ct_dst_sport_ltm	0.21	0.37	0.47	0.14	0.20	0.12	0.26
f37_ct_dst_src_ltm	0.10	0.16	0.38	0.11	0.11	0.12	0.14
f38_is_ftp_login	0.00	0.00	0.01	0.00	0.00	0.00	0.00
f39_ct_ftp_cmd	0.00	0.00	0.01	0.00	0.00	0.00	0.00
f40_ct_flw_http_mthd	0.00	0.01	0.01	0.00	0.00	0.00	0.00
f41_ct_src_ltm	0.09	0.16	0.32	0.09	0.09	0.00	0.12
f42_ct_srv_dst	0.08	0.14	0.32	0.10	0.09	0.13	0.11
f43_is_sm_ips_ports	0.02	0.31	0.20	0.03	0.03	0.15	0.07
f44_attack_cat	1.00	0.68	1.00	1.00	1.00	0.73	1.00

Algorithm 2: Ensemble of kNN using MetaCost **(EKMC)**

Input: subset *F* after applying **PCA**
Output: Performance Metrics
1. Subject input to an Ensemble of MetaCost with kNN Base Classifier.
2. Perform cross validation of ten-folds and record the performance metrics.

Encouraged by the results of kNN, we tried ensembling kNN using Bagging, Classification by Regression and MetaCost and the results as indicated in Table 4 prove that MetaCost happens to be the most efficient of them all.

4. Experimental Results and Discussion

The experiments are carried out on two benchmark data sets UNSW-NB15 and NSL-KDD. The NSL-KDD has 125973 instances in the training and 22544 instances in the test set. EKMC was trained using the training set and was tested making use of the test set. Performance metrics after a more rigorous ten-fold cross validation were then recorded. The performance of the proposed model was later validated using the UNSW-NB15. The UNSW-NB15 comprises of 175,341 instances in the training set and 82,332 instances in the test set. A ten-fold cross validation typically involves dividing the input data set into ten parts and training the model with the nine parts while using the excluded part as the test set and repeating the process for a total of ten times by using an unused test set during each round.

The SDMR Feature Selection algorithm as listed in Algorithm 1 involves the computation of ranks for each feature. Information Gain, Information Gain Ratio, Weight by Correlation, Weight by Chi Squared Statistics, Gini Index, Weight by Tree importance, and Weight by Uncertainty are used for the computation of Ranks. The weights of each feature of NSL KDD data set are listed in Table 1 and that of UNSW-NB 15 in Table 2. The mean value of Weights of Ranks of each Feature as determined by all the chosen Algorithms is initially determined. A Standard Deviation of Mean of Ranks is then Computed. All those Features whose Mean of Ranks is less than or equal to the computed SDMR are dropped and only the Features whose Mean of Ranks is greater than the SDMR are selected. The SDMR of NSL KDD is found to be 0.278831 and that of UNSW-NB 15 is 0.184325. After dropping the Features whose Mean of Ranks is less than the SDMR value, only 15 Features from the NSL-KDD and 11 Features from the UNSW-NB data set are selected.

In addition to SDMR, the data set is also subjected to PCA for dimensionality reduction during the preprocessing stage. The Principal Component Analysis (PCA) when applied on a data set having many variables (features) correlated with one another, reduces its dimensionality by only retaining the variation present in it. The existing variables of the data set are transformed to a new set of variables, known as the principal components (or PCs) that are orthogonal such that the correlation between any pair of variables is 0.The resultant set is subjected to various built-in Classifier Algorithms with ten-fold cross validation to measure the performance metrics. The performance of the Classifier Algorithms is evaluated based on the Accuracy, Classification Error, Precision, Recall, F1-Measure and Detection Rates.

Efficiency of classification would be better when a classifier exhibits true positive rates that are maximum and false positive rates that are minimum. In this context, 8 Performance metrics of classification process are defined. Let N_{ben} represent total number of normal or benign samples and N_{anom} the number of anomalous samples in a data set. True Positive (TP) is the number of normal or benign instances classified correctly as normal is denoted as $N_{ben \rightarrow ben}$ and True Negative (TN) is the number of anomalous instances classified correctly as anomalous is denoted as $N_{anom \rightarrow anom.}$ False Positive (FP) is a measure of normal instances misclassified as anomalous is denoted as $N_{ben \rightarrow anom}$ while False Negative (FN) is a measure of anomalous instances misclassified as normal is denoted as $N_{anom \rightarrow ben}$.

The Detection Rate (DR) is the rate of anomalous samples being classified correctly as anomalous.

$$TPR = \frac{N\ ben \rightarrow ben}{(Nben \rightarrow ben + Nanom \rightarrow ben)} X\ 100 \quad (1)$$

False positive rate (FPR) is the rate of normal samples being classified incorrectly as anomalous samples.

$$FPR = \frac{Nben \rightarrow anom}{(Nanom \rightarrow anom + Nben \rightarrow ben)} X\ 100 \quad (2)$$

False Negative Rate (FNR) is the rate of anomalous samples being classified incorrectly as benign samples.

$$FNR = \frac{Nanom \rightarrow ben}{(Nanom \rightarrow ben + Nanom \rightarrow anom)} X\ 100 \quad (3)$$

True Negative Rate (TNR) is the rate of benign samples being classified correctly as benign out of the total available benign samples.

$$TNR = \frac{N\ ben \rightarrow ben}{(N\ ben \rightarrow ben + Nben \rightarrow anom)} X100 \quad (4)$$

Prediction Accuracy (PA) is the total number of anomalous and benign samples that are identified correctly with respect to the total number of all available samples.

$$PA = \frac{(Nanom \rightarrow anom + N\ ben \rightarrow ben)}{(Nanom \rightarrow anom + N\ ben \rightarrow ben + Nben \rightarrow anom + Nanom \rightarrow ben)} X\ 100 \quad (5)$$

Precision is the number of true positives divided by the total number of elements labeled as belonging to the positive class.

$$Precision = \frac{Nanom \rightarrow anom}{(Nanom \rightarrow anom + Nben \rightarrow anom)} X100 \quad (6)$$

Recall is the number of true positives divided by the total number of elements that really belong to the positive class.

$$Recall = \frac{Nanom \rightarrow anom}{(Nanom \rightarrow anom + Nanom \rightarrow ben)} X\ 100 \quad (7)$$

F1-Measure is the harmonic mean of Precision and Recall and is given by:

$$\text{F1-Measure} = 2\frac{Precision * Recall}{Precision + Recall} \quad (8)$$

The experimental results as indicated in Table 3 suggest that the kNN classifier offers best Prediction Accuracy, Precision, Recall, F-1 measure and Detection Accuracy with least classification error value out of the 20 classifier algorithms. This prompted us to use kNN as the Base Classifier in the Ensembled approach. When different Ensembling Schemes such as Bagging, Classification by Regression and MetaCost were used, only MetaCost with kNN as the Base Classifier offered best results in comparison with the other two approaches as indicated in Table 4 and plotted on a graph as depicted in Figure 3. This was the reason behind choosing MetaCost with kNN as the Base Classifier in our proposed work.

The proposed EKMC performs better than our previous model i.e. ERCRTV [1] and the existing GAA-ADS [8] models when tested on both the data sets. EKMC exhibits good Prediction Accuracy and a better Detection Rate as listed in Table 5 and depicted in Figure 2.

Table 3: Classification result

Classification Algorithm	Accuracy	Classification Error	Precision	Recall	F1 Measure	Detection rate
Perceptron	49.37%	50.63%	68.08%	2.88%	5.26%	2.88%
Neural Net	50.75%	49.25%	48.12%	30.00%	36.96%	30.00%
Quadratic Discriminant Analysis	51.29%	48.71%	4.57%	0.06%	0.12%	0.06%
Regularized Discriminant Model	51.63%	48.37%	8.01%	0.05%	0.10%	0.05%
Naïve Bayes	51.71%	48.29%	21.10%	0.13%	0.26%	0.13%
CHAID	51.88%	48.12%	unknown	0.00%	unknown	0.00%
Default Model	51.88%	48.12%	unknown	0.00%	unknown	0.00%
Generalised Linear Model	51.89%	48.11%	90.91%	0.01%	0.03%	0.01%
Linear Discriminant Model	51.89%	48.11%	90.91%	0.01%	0.03%	0.01%
Linear Regression	51.89%	48.11%	87.50%	0.01%	0.02%	0.01%
Logistic Regression	51.89%	48.11%	68.42%	0.02%	0.04%	0.02%
Deep Learning	52.71%	47.29%	53.91%	88.42%	60.83%	88.42%
Naïve Bayes Kernel	60.32%	39.68%	39.68%	77.63%	77.63%	77.63%
Decision Stump	90.35%	9.65%	92.36%	87.15%	89.68%	87.15%
Gradient Boosted Tree	93.54%	6.46%	91.85%	95.09%	93.42%	95.09%
Random Tree	94.24%	5.76%	92.90%	95.31%	94.09%	95.31%
Rule Induction	94.45%	5.55%	93.96%	94.54%	94.25%	94.54%
Decision Tree	95.54%	4.46%	94.36%	96.51%	95.42%	96.51%
Random Forest	95.74%	4.26%	94.67%	96.58%	95.62%	96.58%
kNN	98.85%	1.15%	98.88%	98.73%	98.81%	98.73%

Table 4: Comparison of Ensembling Techniques

Ensembling Technique	Accuracy	Classification Error	Precision	Recall	F1 Measure	Detection rate
Bagging	98.87%	1.13%	98.88%	98.76%	98.82%	98.76%
Classification by Regression	98.85%	1.15%	98.88%	98.73%	98.81%	98.73%
Meta cost	98.90%	1.10%	98.92%	98.80%	98.86%	98.80%

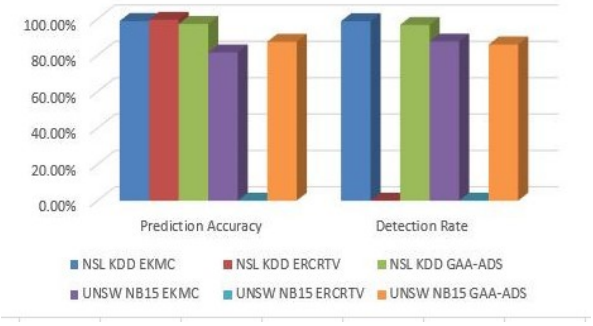


Figure 2: Comparison graph of Various available Models

Table 5: Performance Comparison of the techniques on NSL-KDD and UNSW-NB 15 data sets.

Data set	Method	Accuracy	Detection Rate
NSL KDD	EKMC	98.90%	98.80%
	ERCRTV[1]	99.60%	0
	GAA-ADS[8]	97.30%	96.76%
UNSW NB15	EKMC	81.58%	87.60%
	ERCRTV[1]	0	0
	GAA-ADS[8]	87.46%	86.04%

5. Conclusion

In the Pre-processing phase[9,10], Feature Selection using SDMR is applied to select only significant features from the data set. The *SDMR* Feature Selection algorithm is very much novel and greatly reduces the dimensionality of the data set almost equaling to 70%. It selects only 15 features out of 41 features in case of NSL-KDD and a mere 11 features out of 44 features in case of UNSW-NB15. PCA is then applied to further reduce dimensionality of the data set. The proposed *EKMC* outperforms *GAA-ADS* in terms of Detection rate on both the data sets. The detection rates of EKMC are 98.8% and 87.60% on NSL-KDD and UNSW-NB15 respectively while that of *GAA-ADS* are 96.76 and 86.04% respectively on the same data sets. The performance metrics are recorded for tenfold cross validation. The proposed model is required to be tested on other data sets as well and Classification Error rate must further be reduced.

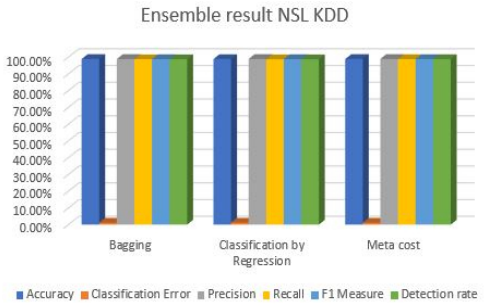


Figure 3: Comparison of Ensembling techniques graph

References

[1] A. Niranjan, D. H. Nutan, A. Nitish, P. D. Shenoy and K. R. Venugopal, "ERCR TV: Ensemble of Random Committee and Random Tree for Efficient Anomaly Classification Using Voting," 2018 3rd International Conference for Convergence in Technology (I2CT), Pune, 2018, pp. 1-5. doi: 10.1109/I2CT.2018.8529797H Haddad Pajouh, R Javidan, R Khayami, D Ali and K K R Choo, "A Two-layer Dimension Reduction and Two-tier Classification Model for Anomaly-Based Intrusion Detection in IoT Backbone Networks", *IEEE Transactions on Emerging Topics in Computing*, vol. pp, no. 99, pp. 1- 11.

[3] Jakub Breier and Jana Branisov,"Anomaly Detection from Log files using Data Mining Techniques", *Proceedings of Conference on Information Science and Applications (Springer),* vol. 339, 2015, pp. 449-457.

[4] Bayu Adhi Tama, Akash Suresh Patil, Kyung-Hyune Rhee "An Improved Model of Anomaly Detection using Two-level Classifier Ensemble", *Proceedings of 12th Asia Joint Conference on Information Security,* 2017.

[5] Ziyu Wang, Jiahai Yang, Zhang ShiZe, Chenxi Li "Robust Regression for Anomaly Detection", *Proceedings of IEEE ICC 2017 Communication and Information Systems Security Symposium,* 2017.

[6] Teodora Sandra Buda1, Haytham Assem1 and Lei Xu1 "ADE: An Ensemble Approach for Early Anomaly Detection", *Integrated Network and Service Management (IM), 2017 IFIP/IEEE Symposium,* 2017.

[7] A. Niranjan, K. M. Akshobhya, P. D. Shenoy and K. R. Venugopal, "EKNIS: Ensemble of KNN, Naïve Bayes Kernel and ID3 for Efficient Botnet Classification Using Stacking," *2018 International Conference on Data Science and Engineering (ICDSE)*, Kochi, 2018, pp. 1-6.doi: 10.1109/ICDSE.2018.8527791

[8] N. Moustafa, J. Slay and G. Creech, "Novel Geometric Area Analysis Technique for Anomaly Detection using Trapezoidal Area Estimation on Large-Scale Networks," in *IEEE Transactions on Big Data*. doi: 10.1109/TBDATA.2017.2715166

[9] Niranjan A, Nitish A, P Deepa Shenoy and Venugopal K R, "Security in Data Mining-a Comprehensive Survey", *Global Journal of Computer Science and Technology*, vol. 16, no. 5, 2017, pp. 52-73.

[10] Asha S Manek, Samhitha M R, Shruthy S , Veena H Bhat, P Deepa Shenoy, M. Chandra Mohan, Venugopal K R, L M Patnaik "RePID-OK: Spam Detection using Repetitive Pre-processing", *IEEE CUBE 2013 Conference, ISBN : 978-1-4799-2234-5*, pp. 144-149, November 15-16, 2013.

ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 5, 409-418 (2019)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

A Big Data Security Layer Meta-Model Proposition

Allae Erraissi*, Abdessamad Belangour

Laboratory of Information Technology and Modeling, Hassan II University, Faculty of sciences Ben M'Sik, Casablanca, Morocc.

ARTICLE INFO

Article history:
Received: 11 July, 2019
Accepted: 09 October, 2019
Online: 28 October, 2019

Keywords:
Meta-model
Model Driven Engineering
Big Data
Security layer
Management layer

ABSTRACT

In Big Data, several solution providers offer distributions to handle this large amount of data. Given the variety of these solutions, we are working to provide universal meta-modeling for all layers of a Big Data system in order to address the issue of interoperability and portability between these solutions. As part of our continuous efforts to standardize concepts in Big Data world, we apply in this paper techniques related to Model-Driven Engineering "MDE" to propose a meta-model for the security layer in Big Data. This meta-model with the others that we have already proposed for the other layers of the Big Data system, will be used as a platform-independent according to Model-Driven Architecture pattern, which describes the structures of Big Data layers independently of any specific platform.

1. Introduction

The issues of the expression and application of security needs are present in all information systems. With the emergence of new types of environments, such as cloud computing and big data, this problem is becoming more complex. Hence, it is necessary to take into account their heterogeneity and the different levels of such an architecture to deal with this problem.

At the Big Data level, several distribution providers have proposed solutions to manage this huge amount of data, among these distributions, we found HortonWorks, Cloudera, MapR, InfoSphere BigInsghits IBM, Pivotal HD, etc. To put it another way, this massive amount of heterogeneous data led to the emergence of a large number of big data systems and technologies that share similar architectures but with different implementations. In essence, the common architecture of these data systems is composed of many components: Data sources, Ingestion, Hadoop Storage, Hadoop Platform management, Visualization, Monitoring, and Security Layers [1]. In our way for a unified abstract implementation, we proposed, in previous works, meta-models for data sources, ingestion [2], storage [3], management [4] and visualization layers [5]. We also relied on our previous comparatives studies to define key concepts of security layer in Big Data [6].

In this paper, we shall first present the security properties that express the security requirements of the software architecture. These properties are then combined to form a security policy: different models and languages of expression of security and insurance policies will be detailed. Then, we shall present the meta-model that we proposed for the security layer. This meta-model with the others already proposed for the other layers of the Big Data system will be used like a platform-independent according to Model-Driven Architecture pattern [7], which describes the structures of Big Data layers independently of any specific platform.

2. Related Work

Several research studies have been developed to standardize the security layer concepts at the Big Data systems level. Among the foremost relevant approaches in security modeling are UMLSec [8] and SecureUML [9]. Yet, both approaches use UML extensions to specify security requirements. UMLSec was created to verify the formal security requirements specifications within the system design. SecureUML was used to illustrate the MDS approach. SecureUML is specific to role-based access control infrastructures. It shows how an MDE policy could be applied to code generation for any aspect of security.

In the literature, several MDS approaches have been proposed. For example, on a middleware platform that integrates the implementation of the Corba component model with the OpenPMF security framework, Reznik et al. [10] present an MDS solution for

*Allae Erraissi, Laboratory of Information Technology and Modeling, Hassan II University, Faculty of sciences Ben M'Sik, Casablanca, Morocco, Email: erraissi.allae@gmail.com

developing secure applications. In the MDS approach, another UML profile was developed to model access control policies. Lang and Schreiner [11] show how OpenPMF architecture can be used to translate a high-level security-related regulatory requirement into enforceable authorization rules.

Our proposal is different from other existing MDS approaches, which use templates to represent design decisions and implementation details for target platforms. The application of Model Driven Engineering techniques that provide meta-modeling to the security layer in Big Data will advocate a systematic MDS application process. This application separates security requirements specifications and design decisions related to their implementation. Thus, the reuse of models is facilitated by this separation.

However, ModelSec proposes an intermediate model conforming to a target platform meta-model to reduce the semantic gap between the security requirements and the software generated. This intermediate step favors the reuse of the transformations. Unlike the other approaches that create UML profiles, a DSL is defined to express the security requirements. It has been implemented by applying meta-modeling techniques. Besides, ModelSec is a generic approach, whereas most MDS approaches are specific and they focus on access control policies.

Yet, other approaches manage security requirements. These approaches are not aligned with MDS, and they focus primarily on eliciting security requirements rather than generating software objects from them. Among these proposals are Secure Tropos [12] and [13], which introduce the concept of antigoal. Yu et al. [14] use ontologies and Haley et al. [15] present a framework consisting of a set of activities designed to meet security requirements.

3. Security Properties and Policies

In this section, we describe methods for formalizing the different security needs. We first present the usual security properties and those that can be derived from them. Then, we shall take existing models to apply these properties, security policies and languages to define them.

3.1. Security properties

Security properties are the basis for expressing security requirements. The set of security properties is commonly seen as a set derived from three main properties: confidentiality, integrity, and availability (CIA: Confidentiality, Integrity, Availability). The exact interpretation of what these three properties imply depends on the context of use. However, their definition and application is an essential part of the safety assessment criteria, at both European [16] and international [17]. Several definitions of these properties exist in the literature [16,17,18]. We present a synthesis here.

3.1.1. Confidentiality

Confidentiality is about preventing unauthorized disclosure of information. It aims to prohibit unauthorized access to information. Accordingly, the property of confidentiality implies that the information is accessible only by certain entities, that is to say, that certain entities must not be able to obtain the information. This property is often used in sensitive environments, such as defense. It concerns both direct access and information transfer.

3.1.2. Integrity

Integrity means the fact that information cannot be unauthorized modified or deleted, whether during processing, storage or transfer of information. This property does not only concern voluntary modifications and deletions, but also accidental acts. Just as in the case of confidentiality, the integrity property specifies all the entities authorized to modify or delete information. By default, other entities cannot alter the information.

3.1.3. Availability

The availability property expresses the ability to access information or a resource. It is related to the reliability of a system since an unavailable system is a failing system. The temporal notion of access is relative to the field of application: access to information or service on a critical system (for example, in the medical field) must be done more quickly than on a non-critical system (for example, a website).

3.1.4. Insurance

The property triple of confidentiality, integrity, and availability is sometimes extended with other properties [19], such as the insurance property. The purpose of this is to provide evidence that the other properties have been applied and that they have the desired effect. The assurance of a property is the verification that this property has been properly applied. The level of insurance required depends on the system in question and must, therefore, be adapted according to the criticality of the system. In sum, the purpose of insurance property is to verify that the other properties of the policy have been applied and that this application has had the desired effect.

3.1.5. Derived properties

Privacy, integrity, and availability properties are the basic concepts of security. They can be used to define derived properties which are special cases, subsets or combinations of these basic properties. In this section, we describe some of these derived properties.

- PROCESS CONFINEMENT

Process containment has been defined by Lampson [20]. The problem of containment concerns the prevention of the disclosure by a service or a process of information considered as confidential by the users of this service. According to Lampson, one of the features needed for a process is that it should not disclose information and it must not store information. Indeed, if a process stores information and a user can observe this process, then there is a risk that the user can access the information. If the process does not store the information, then it cannot be disclosed. This process containment property can thus be seen as the isolation of the process from the rest of the system.

- AUTHENTICATION

The authentication property is a property that allows or denies access to information or service to entities. Authentication is the process of establishing trust in the identity of an entity [21]. This authentication property is essential to apply the properties seen previously. Indeed, it makes it possible to establish the identity of

a user, a service or a system, which is necessary to give him the appropriate rights.

3.2. Security Policies

A security policy is a set of properties that express the security needs of a system or a set of systems. Most security properties are based on resource access control. As a result, the concepts introduced in access control models can be transposed into a language to formally express the properties of the previous section. In this section, we shall first describe the main models of historical access control. Then, we shall present security policy and insurance policy expression languages, which express and apply security properties.

3.2.1. Historical models

As defined above, some security properties can be applied by access control mechanisms. An access control system is usually modeled using the following three elements:

- A set of topics that are the active entities of the system (for example, processes);
- A set of objects that are the passive entities of the system, on which the subjects can perform actions (files, sockets, etc.);
- A set of permissions that represent the authorized actions between a subject and an object (reading, writing, etc.), or between two subjects (sending a signal).

- **Control of Discretionary Access**

Discretionary Access Control (DAC) is the default historical model present on most operating systems. In this model, the management of access rights to a resource is left to the discretion of the owner of that resource. For example, on UNIX, the owner of a file can set the read, write, and execute rights for itself, for the members of the group that owns the file, and for all other users in the system.

An access control model can be represented as a matrix, where a line represents a subject, a column represents an object or a subject, and each element of the matrix represents a set of permissions of the subject on the object (or on the second subject). This model was formalized by Lampson [22, 23] using capacity lists and Access Control Lists (ACLs). Therefore, it proposes to indicate, in an array A, the set D of the protection domains (representing program execution contexts, that is to say, the subjects) on the lines, and the set X of the objects on the columns. Hence, Lampson defines the lists of capacities (definition 1) which establish the permissions of a domain d on all the objects o of the system. This is the set of actions allowed for each domain d.

Definition 1: List of Capabilities

Given a d∈D domain, the list of capabilities for domain d is all couples:

$$\mathbf{(o, A[d, o]), \forall o \in X.} \quad (1)$$

Afterworlds, Lampson defines access control lists (definition 2) that specify the set of permissions granted on an object for each domain in the system.

Definition 2: Access Control List (ACL)

Given an o∈X object, the access control list (ACL) for object o is the set of pairs:

$$\mathbf{(d, A[d, o]), \forall d \in D.} \quad (2)$$

However, this model is complex to update. For example, when adding a new user to the system, a complete row must be added to the A matrix. The Lampson model has therefore evolved to the HRU model. In the HRU model [24], discretionary access control is modeled from a matrix P containing all the subjects' rights to the objects. In this template, subjects can edit the access control matrix to create or delete topics or objects, as well as edit existing permissions. The HRU model models protection using the following elements:

- An access control matrix P;
- A set S of subjects and a set O of objects;
- A set R of generic rights (reading, writing, execution, possession, etc.);
- A finite set C of commands representing all the operations provided by the operating system (creation of files, modification of rights, etc.);
- A set E of elementary operations: enter and delete (add and remove rights), create subject and create object, destroy subject and destroy object (destruction of subjects and objects).

A triplet (S, O, P) represents the configuration of the system protection. In order to study the problem of the safety of a protection system, the authors of the model HRU are interested in the transfer of privilege (right) occurring when a command inserts a right r in the matrix P. The safety problem can be defined in the following way: given an initial configuration of the security policy, a system is considered safe for a right r if none of the commands of this system causes the transfer of the right r. The authors then showed that if the commands contain only one elementary action, the safety problem is decidable but its verification algorithm is NP-complete. Nevertheless, in the general case (the commands contain several elementary actions), the problem is undecidable. In the case of an operating system, the commands are not mono-operational and the problem of the safety of the protection system is therefore undecidable. This shows that it is impossible to guarantee security properties with a discretionary access control model.

Other DAC models have been defined to extend the HRU model, including TAM (Typed Access Matrix) [25] and DTAM (Dynamic Typed Access Matrix) [26]. TAM extends the HRU model by incorporating a strong typing concept [27] which corresponds to the association of immutable security types to all the subjects and objects of the system. DTAM extends the TAM model by adding the ability to dynamically modify object types. The different models of discretionary access control are historical models. They are mainly used for system rights management.

- **Mandatory access control**

Discretionary access control leaves the management of resource permissions to their owners, that is, users of the system.

In practice, this system is limited since many attacks against systems aim to obtain privileged access (root, or super-user). Such an attack, so-called elevation of privilege, is intended to obtain rights greater than those possessed. Thus, it allows the user to gain full access to the system and its resources and to overcome discretionary control. Besides, studies [28,29] have shown that DAC models are vulnerable, particularly because of the need for users to correctly define the set of resource permissions. Any definition error can create a security vulnerability that can be exploited to gain privileges. Mandatory Access Control (MAC) is consequently intended to address this problem by imposing a security policy on system users. Anderson proposes the use of a reference monitor [30] to control the interactions between subjects and objects and to determine which ones are valid (that is, the ones allowed by the policy). This section shall present the main models of mandatory access control. These models explain the concepts of security properties (integrity, confidentiality) in order to apply them.

The Bell-LaPadula (BLP) model [31] is based on the confidentiality needs of the military and is accordingly intended to prevent disclosures. This model extends the HRU model by adding the notion of a label associated with each subject and object of the system. A label corresponds to a security level and is composed of two security identifiers: the first identifier, hierarchical, indicates the level of classification (for objects) or authorization (for subjects), e.g., secret or top secret. The second identifier, called category, specifies the organizations using the information, for example, military or private.

In addition to the classic rules which are defined by an access control matrix, two new rules are defined:

- **ss-property** (simple security property): for reading access to be authorized, the subject requesting it must have a level of authorization greater than or equal to that of the object;
- ***-property** (star property): Information can only be transferred from a lower classification object to a higher classification object.

These two rules make it possible to ensure the confidentiality of the information and its non-disclosure. However, the existence of hidden channels can cause information flows that cannot be controlled. For this reason, a more restrictive version of BLP has been proposed with the following rules:

- No Read Up: A subject requesting read access to an object must have a security level greater than or equal to the object;
- No Write Down: A subject requesting write-only access (adding data) to an object must have a security level that is less than or equal to the object.

Consequently, a subject requesting read and write access to an object must have the same level of security as the object. This model is sometimes called MLS (Multi-Level Security), and thus refers to the level system used to define security rules.

A dual model at BLP, Biba [32] has been defined to meet integrity needs. The Domain and Type Enforcement (DTE) model [33] is a high-level mandatory access control model. Unlike models such as BLP or Biba, it is not intended to apply a specific security property, but rather to define the allowed access between different entities in the system. Nonetheless, this model can serve as a basis for implementing security property models, such as BLP, Biba, or process containment. The DTE replaces notions of topics and objects with those of domains and types. Thus, each object has a type and each subject runs in a domain. Access rights on types and domains are then defined for each domain. Consequently, this model aims to restrict the resources accessible by a process (even for privileged processes by the principle of least privilege) and to control which processes have access to sensitive resources. The different models of mandatory access control can be applied at the same time at the system level, in distributed environments such as the case of military infrastructures, or even be integrated within a software.

3.2.2. Expression languages of security policies

The access control models that were presented above are the historical models that allowed the definition of security policy expression models and languages. A security policy consists of a set of rules that may have the purpose of applying properties, possibly following one of the described models or reusing certain concepts. In this section, we shall describe the main models and languages of a policy expression.

RBAC [Role Based Access Control] is a role-based access control model that simplifies the writing and management of a security policy. The administration of a mandatory access control policy involves the management of multiple access rules. Yet, this process is time-consuming and subject to errors. With RBAC, rules can be simplified by expressing them according to the role of the subject and not his or her identity.

Thus, within an organization, roles are defined from the functions of the different positions and permissions are assigned to these roles. Users have roles that allow them to obtain permission. Considering that the roles are not directly assigned to the users, their management is facilitated. For example, when adding a user, it is necessary to assign the corresponding roles.

Different versions of RBAC have been defined [34,35]:

- Core RBAC understands the basic concepts of RBAC and specifies that user-role and role-permission associations are of many-to-many types;
- Hierarchical RBAC adds inheritance support between roles: for example, a junior employee has a junior role with permissions and the senior role inherits permissions from the junior role;
- Constrained RBAC introduces constraints to apply the privilege separation principle. Since 2004, RBAC is a NIST standard [36].

OrBAC [37] (Organization Based Access Control) is an access control model based on the concept of organization. OrBAC abstracts the notions of subject, action, and object by those of role (as in the case of RBAC), activity and view. Subsequently, this makes it possible to group entities according to the security rules that concern them, and thus simplify the expression of the policy. In order to facilitate the expression of dynamic policies, OrBAC uses contexts. Three types of rules are possible: First, permissions which allow a role to perform an activity on a view, in a given

context. Second, bans when an activity is prohibited and finally, obligations when a role must perform an activity on a view in a given context. An OrBAC rule can be expressed as follows: A role can have permission, prohibition or obligation to perform an activity on a given view when the associated context is verified.

ABAC [38] (Attribute Based Access Control) is an access control model in which rights are given to users based on attributes. Attributes can be seen as features of system elements. An attribute consists of a type (for example, a role, a project, a sensitivity level, etc.) and a value that can be single-valued or multi-valued. For example, a role identifier or the value of the sensitivity level, etc.). These attributes can be compared to each other or to fixed values, which allows defining the rules of policy. For example, a rule evaluates one or more attributes of a subject to decide whether to allow or deny access to an object. Note that ABAC can be used with the DAC, MAC and RBAC models [39].

XACML [40] (eXtensible Access Control Markup Language) is a standard of the OASIS consortium [41]. It is both an access control policy definition language and a template for interpreting (allowing or disallowing) access requests based on this policy. XACML is based on the XML language. A set of XACML policies consists of policies, which are themselves composed of a set of elements: target, rules, a combination algorithm, and obligations. The target element indicates whether a policy should be applied to a given query. The rules specify the allowed accesses: they consist of a condition that is to say a Boolean expression. If this condition is verified, it causes an effect (authorization or prohibition) on the access. The rule combination algorithm defines how the policy should be interpreted when multiple rules may apply, for example, "permission overrides prohibition". Obligations are optional and allow you to perform an action when a rule is encountered, for instance, generate an alert. However, because of the flexibility and expressiveness of XACML, the definition of security policies can be complex [42]. XACML cannot easily allow the cloud user to express his own security needs.

Ponder [43] is a specification language for security and administration policies for distributed systems. It is a declarative and object-oriented language that supports different types of policies: authorization policies (specifying allowed or forbidden access for a subject), obligation (actions that a subject must perform in response to an event), restriction (actions that a subject should not perform), delegation (actions that a subject may delegate to another subject), constraint (to limit the application of other policies, for example in function of time). Meta-policies can also be defined to manage interactions between policies. Ponder2 [44] reuses Ponder's concepts by adapting them to autonomous systems. Hence, Ponder2 is based on a decentralized architecture, made up of self-managed components (SMCs) capable of interpreting and applying the policy on system objects. The objects considered by Ponder2 are Java objects that can communicate with SMCs. These objects must be adapted to be managed by Ponder2.

3.2.3. Expression language of insurance policies

Security policies may be supplemented by insurance policies. An insurance policy must express methods for assessing the level of protection of a system.

XCCDF [45] is an XML-based standard for specifying security checks and performance tests. Hence, it is, a language used for defining an insurance and verification policy. Besides, XCCDF can be used to automate vulnerability checks and responses when these vulnerabilities are detected. XCCDF is used by SCAP [46] (Security Content Automation Protocol), which is a set of specifications defined by NIST to standardize the format in which software vulnerabilities and security configurations are expressed. A distributed version of XCCDF, called DXCCDF [47], is used to express distributed vulnerabilities.

OVAL [48] is another standard aimed at unifying the expression of insurance policies. OVAL uses XML to evaluate the state of a system by performing a series of tests on the machine. The evaluation of the state of the system with OVAL is broken down into three stages: the representation of the system information, the description of the different states and the transmission of the results of the evaluation. The distributed version of OVAL is named DOVAL [49] and can handle distributed vulnerabilities.

A-PPL [50] is a policy language for expressing accountability obligations. A-PPL can be used for privacy management policies, access control, and usage control policies. Besides, A-PPL manages elements specific to the insurance: it makes it possible to define alerts in the event of a detected error and determine rules of localization of the data. It can also specify which are the characteristics wanted for the audit and the storing error messages.

4. Meta-model for Security layer:

Big Data solutions are more and more used. Indeed, many providers of Big Data solutions have already proposed distributions like HortonWorks, Cloudera, MapR, etc. Although each of these distributions has its vision for a Big Data system, they all share their necessary needs of the security layer. Accordingly, Security is a crucial element for both customers and service providers.

However, before presenting the Security layer, we deem it necessary to talk firstly about our latest contribution of meta-models. In fact, in our previous work, we have proposed meta-models for layers: Data Sources, Ingestion, Hadoop Storage [51], Hadoop Platform Management [52], and visualization of Big Data architecture. Accordingly, we realized that there is a direct relationship between the security layer and the other layers in Big Data. In order to show the link between the layers of the Big Data system, we shall present in figure 1 the following meta-package diagram. It clearly indicates the meta-packages: IngestionPkg, DataSourcesPkg, HadoopPlatformManagementPkg, VizualisationPkg, SecurityPkg, and MonitoringPkg and the dependency relationship that links them.

We present in the figure 2 the meta-model that we proposed for the Security layer in Big Data. This meta-model defines specific security concepts. To separate the access control mechanisms from the rest of the concepts, we have devised our meta-model to two separate parts: the meta-model of security concepts and the meta-model that defines access control mechanisms. The purpose of this separation is to offer the possibility of extending the meta-model of access control mechanisms in our future work. It should be mentioned that this extension of the meta-model will not affect the rest of the meta-model. The figure 2 shows the meta-model of the general concepts of the security layer.

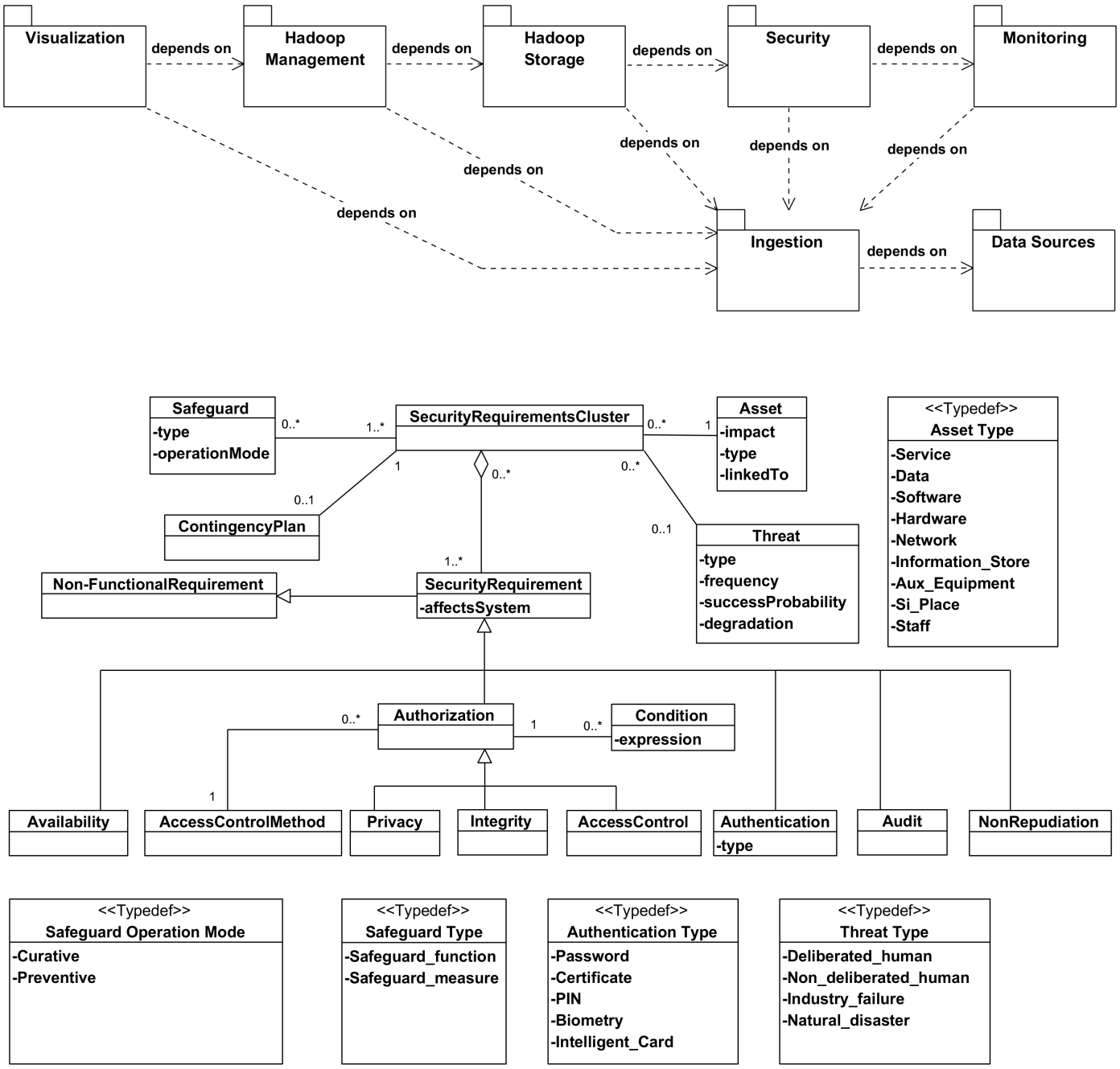


Figure. 2. Meta-model of general security concepts.

A threat can damage the assets. A threat has the following properties: type, frequency (modeled as an annual rate), probability of actual success, and degradation (that is, the level of damage to an asset if a threat reaches its goal).). Safeguards are a barrier to risk to reduce it. As indicated in the Safeguards type attribute, we can distinguish between Safeguard functions and Safeguard measures. Safeguards functions are actions that reduce risk while Safeguards' measures are physical or logical devices or

processes that reduce risk. Safeguards can act in two different modes, healing if they act on damaged assets or preventative if they act before a threat has appeared. A detailed emergency plan which consists of a set of safeguards is recommended to reduce a threat that may cause harm.

There is no standard classification for security requirements. According to [53], seven categories were taken into account (confidentiality, access control, integrity, authentication, non-repudiation, availability, and audit). Integrity is the guarantee that the information remains complete and correct. Access control is employed to force authorized users to access an asset. Confidentiality is to ensure that only authorized information can be read by those who are authorized. These three types of requirements are related to authentication and may have a condition that is defined as an expression. Authentication is a procedure by which a computer system certifies the identity of a person or a computer. The purpose of this procedure is to allow the person to access certain secure resources. It will compare the information of authorized users stored in a database (locally or on an authentication server) to those provided. Access will only be allowed if the information is the same. It is the administrator of the information system who grants the rights and sets the access. The user with an access account (ID + password) will only have access to the resources he is allowed to see. Availability maintains the proper functioning of the information system and ensures access to a service or resources. Finally, the audit assesses the IT risks of physical security, logical security, change management, emergency plan, etc. It also assesses a set of IT processes - which is usually the case - to respond to a specific customer request. All of these kinds of requirements can affect particular assets or the entire system.

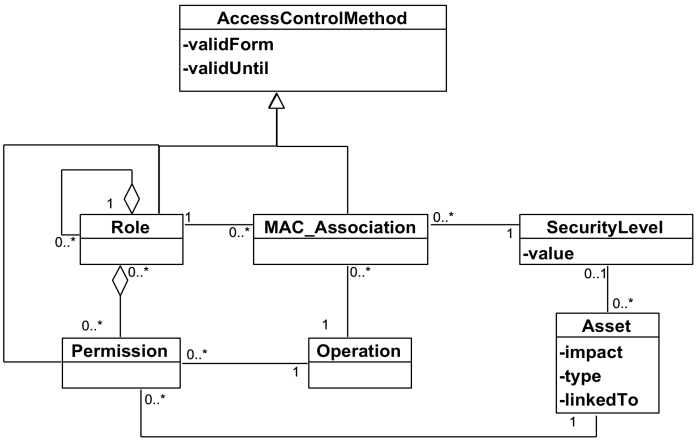


Figure. 3. Meta-model of access control mechanisms.

These two meta-models that we have proposed show the most important attributes, concepts, and data types that a Big Data system would need. Indeed, the meta-class of non-functional requirements is the extension point of the main meta-model and has the ability to add new non-functional requirements. The security requirement is a non-functional requirement specialization intended to be the root of meta-classes representing security concepts.

5. Transformations

The meta-models that we have defined for the different Big Data layers represent the platform-independent model (PIM) level, which means a model-independent of the platform. The goal of these meta-models is to standardize concepts at the Big Data level and to create an independent meta-modeling of platforms and solutions. Following that, we made some transformations using the Atlas Transformation Language (ATL). This transformation language will allow us to move from the PIM model to the Platform Specific Model (PSM). PSMs can use domain-specific languages or general languages like Java, C #, Python, and so on. The techniques used in the MDA approach are thus mainly modeling techniques and model transformation techniques. The figure below shows the transformations that we have made in this paper to make the transition from our meta-model of the security layer at the level of Big Data to PSM:

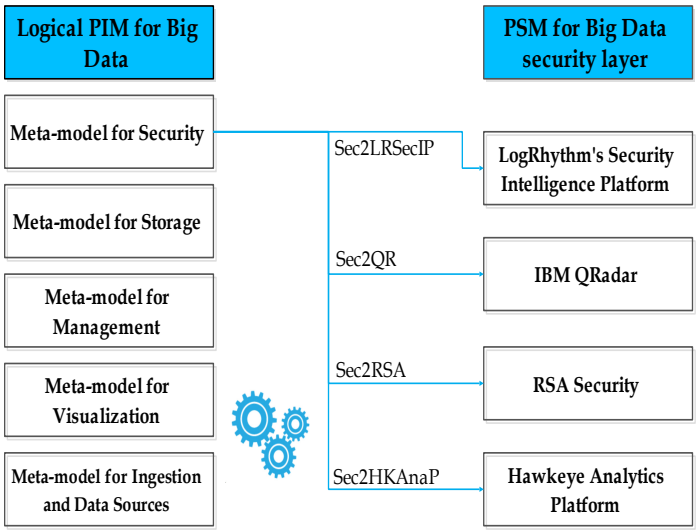


Figure. 4. Transformation of security meta-model to PSM.

6.1. Configuration

In this section, we shall discuss the techniques that we have used to implement the approach presented in Figure 4. Version 4.12 of the Eclipse IDE was used, with the addition of the Eclipse Modeling Framework (EMF) to draw the proposed meta-models. We also used version 4.1 of the ATL transformation language on

Eclipse 4.12, to define the transformation rules and to create our four transformations: Sec2LRSecIP, Sec2QR, Sec2RSA, and Sec2HKAnaP. The switch to the PSM did not take into consideration the version of the chosen solutions since the meta-model we proposed is a standard for the security layer. After applying the transformation rules, the result found will be used on all versions of LogRhythm's Security Intelligence Platform, IBM QRadar, RSA Security, and Hawkeye Analytics Platform.

6.2. Experiences

While performing our tests, we used two datasets to better measure the transformation execution time to the four chosen solutions: LogRhythm's Security Intelligence Platform, IBM QRadar, RSA Security, and Hawkeye Analytics Platform. The results are shown in Figures 5 and 6.

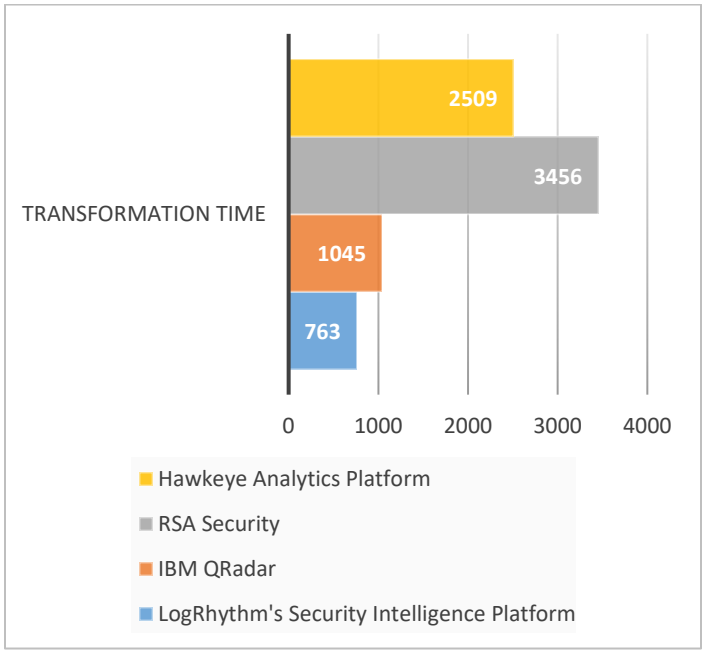


Figure. 5. Transformations time.

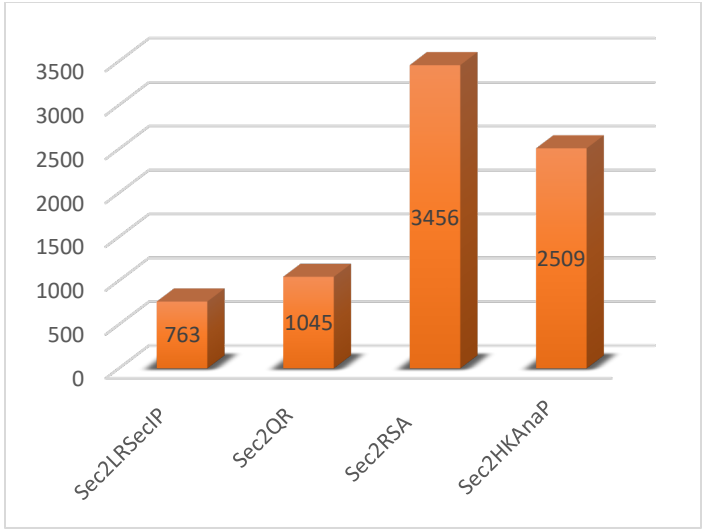


Figure. 6. Execution time for ATL transformations.

Table 1: Transformations runtime.

Security solution	LogRhythm's Security Intelligence Platform	IBM QRadar	RSA Security	Hawkeye Analytics Platform
Transformation time	763	1045	3456	2509

7. Discussion and Perspectives

The security properties that were presented in the first section are textual and abstract. They allow a user literally to express his security needs. Although these properties are high-level requirements, they cannot be directly applied to the system. This is why different models and security languages are designed to express these properties so that they can be applied to a system.

In the second section, we presented historical models for access control policies and security and insurance policy expression languages. Indeed, Historical access control models introduce concepts. These concepts can be generalized to security policies that are not dedicated to access control. For example, modeling in subject-permission-object triplet form is not necessarily specific to access control. Besides, access control has shown the value of a mandatory security policy compared to a discretionary policy. In fact, a discretionary policy gives the possibility to guarantee security properties [56,57]. Existing policy models and languages also make it possible to introduce essential notions into the expression languages of security and insurance policies. Thus, the notion of role introduced in RBAC is used in other models of security policies, such as OrBAC and XACML. It can simplify the expression and administration of policy. These languages allow the definition of security needs in the form of properties. However, these policies are often restricted to the expression of a property and support only a subset of properties. Moreover, existing languages are mostly not suitable for defining security and assurance properties for cloud and big data environments.

As we have mentioned before, Big Data has become essential in today's world. Indeed, our analysis of the subject leads us to find that there are several solutions to manage Big Data in the market (HortonWorks, Pivotal HD, IBM InfoSphere BigInsights, etc.). Each of these solutions manages Big Data in its own way without relying on standards like meta-models. This of course results in the diversity of solutions and the non-interoperability between the different solutions. During our research project, we have worked on the meta-modeling of the different layers of a Big Data system [58]. In this paper, we rely on comparative studies that we have already done to draw the key concepts of the security layer. The meta-model that we proposed for this layer and the other layers is platform-independent according to Model-Driven Architecture pattern. By using the ATL transformation language, for example, we can move to the other platform-specific models (PSM) [59,60]. The PSM we have chosen in this work are LogRhythm's Security Intelligence Platform, IBM QRadar, RSA Security, and Hawkeye Analytics Platform. After the creation of these meta-models, in the next step, we shall work on the creation of models respecting these meta-models. Then we shall define the transformation rules between these meta-models using the transformation language ATL (Atlas Transformation Language). These meta-models are platform-independent according to Model Driven Architecture

pattern, which describes the structures of Big Data layers independently from any specific platform.

8. Conclusion

Big data refers to the collection and aggregation of large amounts of data from different sources to extract information through statistical, descriptive and predictive analysis. Several new solutions exist in the IT market that can handle this huge amount of data. This big data trend makes security an essential but complex point to address. The definition of security needs can indeed be a difficult task, especially since the user of the service does not necessarily know these needs. However, we have seen that there are methods of risk analysis that allow a user to determine their security needs. Thus, we have presented in this paper the different basic security properties that may correspond to the needs of a user, as well as several models of policies to express these needs. After that, we have defined a standard meta-model for the security layer that represents the PIM level. Finally, we made transformations to move from our meta-model to four PSM.

References

[1] N. Sawant and H. (Software engineer) Shah, Big data application architecture Q & A a problem-solution approach. Apress, 2013.

[2] Erraissi, A., & Belangour, A. Data sources and ingestion big data layers: meta-modeling of key concepts and features. International Journal of Engineering & Technology, 7(4), 3607–3612, 2018. https://doi.org/10.14419/ijet.v7i4.21742

[3] Erraissi, Allae, and Abdessamad Belangour. « Hadoop Storage Big Data Layer: Meta-Modeling of Key Concepts and Features ». International Journal of Advanced Trends in Computer Science and Engineering 8, nº 3, 646-53, 2019. https://doi.org/10.30534/ijatcse/2019/49832019

[4] Erraissi, Allae, and Abdessamad Belangour. « Meta-Modeling of Big Data Management Layer ». International Journal of Emerging Trends in Engineering Research 7, no 7, 36-43, 2019. https://doi.org/10.30534/ijeter/2019/01772019.

[5] Erraissi, Allae, and Abdessamad Belangour. « Meta-Modeling of Big Data Visualization Layer Using On-Line Analytical Processing (OLAP) ». International Journal of Advanced Trends in Computer Science and Engineering 8, nº 4, 990-98, 2019. https://doi.org/10.30534/ijatcse/2019/02842019.

[6] Allae Erraissi, Abdessamad Belangour, and Abderrahim Tragha, "Digging into Hadoop-based Big Data Architectures," Int. J. Comput. Sci. Issues IJCSI, vol. 14, no. 6, pp. 52–59, Nov. 2017. https://doi.org/10.20943/01201706.5259

[7] Kleppe, A.G., Warmer, J., Warmer, J.B. and Bast, W., 2003. MDA explained: the model driven architecture: practice and promise. Addison-Wesley Professional, 2003.

[8] Jurjens, J. Secure Systems Development with UML. Springer Verlag, 2003.

[9] Basin, D., Doser, J., and Lodderstedt, T. Model driven security: From UML models to access control infrastructures. ACM Trans. Softw. Eng. Methodol., 15(1):39–91, 2006. https://doi.org/10.1145/1125808.1125810

[10] Reznik, J., Ritter, T., Schreiner, R., and Lang, U. Model driven development of security aspects. ENCTS, 163(2):65–79, 2007. https://doi.org/10.1016/j.entcs.2006.10.016

[11] Lang, U. and Schreiner, R. Managing business compliance using model-driven security management. In Securing Electronic Business Processes, pages 1–11, 2008. https://doi.org/10.1007/978-3-8348-9283-6_24

[12] Bresciani, P., Mouratidis, H., and Zanone, N. Modelling security and trust with secure tropos. Integrating Security and Software Engineering: Advances and Future Visions, pages 160–189, 2007. https://doi.org/10.4018/978-1-59904-147-6.ch008

[13] Van Lamsweerde, A. Elaborating security requirements by construction of intentional anti-models. In ICSE'04: 26th Int. Conf. on Software Engineering, pages 148–157. IEEE Computer Society, 2004. https://doi.org/10.1109/ICSE.2004.1317437

[14] Yu, E., Liu, L., and Mylopoulos, J. A social ontology for integrating security and software engineering. Integrating Security and Software Engineering: Advances and Future Visions, pages 70–109, 2007. https://doi.org/10.4018/9781605660608.ch048

[15] Haley, C., Laney, R., Moffett, J., and Nuseibeh, B. Security requirements engineering: A framework for representation and analysis. IEEE Trans. Software Eng., 34(1):133–153, 2008. https://doi.org/10.1109/TSE.2007.70754

[16] C.Jahl, ITSEC. Information Technology Security Evaluation Criteria (ITSEC) v1.2. Technical report, 1991. https://doi.org/10.1109/ICSE.1991.130656

[17] Tcsec, D. Trusted computer system evaluation criteria. DoD 5200.28-STD, 83, 1985.

[18] Bishop, M. What is computer security? Security & Privacy, IEEE, 1(1):67–69, 2003. https://doi.org/10.1109/MSECP.2003.1176998

[19] Stoneburner, G. Underlying technical models for information technology security: recommendation of the National Institute of Standards and Technology. US Department of Commerce, Computer Security Division, Information Technology, National Institute of Standards and Technology, 2001.

[20] Lampson, B. W. A note on the confinement problem. Communications of the ACM, 16(10):613–615, 1973. https://doi.org/10.1145/362375.362389

[21] Burr, W. E., Dodson, D. F. et Polk, W. T. Electronic authentication guideline: Recommendations of the national institute of standards and technology. NIST Special Publication, pages 800–63, 2013. https://doi.org/10.6028/NIST.SP.800-63-2

[22] Lampson, B. W. Dynamic protection structures. In Proceedings of the November 18-20, 1969, fall joint computer conference, pages 27–38. ACM, 1969. https://doi.org/10.1145/1478559.1478563

[23] Lampson, B. Protection. In Proc. 5th Princeton Conf. on Information Sciences and Systems, pages 18–24. Princeton, 1971. https://doi.org/10.1145/775265.775268

[24] Harrison, M. A., Ruzzo, W. L. et Ullman, J. D. Protection in operating systems. Communications of the ACM, 19(8):461–471, 1976. https://doi.org/10.1145/360303.360333

[25] Sandhu, R. S. The typed access matrix model. In Research in Security and Privacy. Proceedings. 1992 IEEE Computer Society Symposium on, pages 122–136. IEEE, 1992. https://doi.org/10.1109/RISP.1992.213266

[26] Soshi, M., Maekawa, M. et Okamoto, E. The dynamictyped access matrix model and decidability of the safety problem. IEICE transactions on fundamentals of electronics, communications and computer sciences, 87(1):190–203, 2004. https://doi.org/10.1007/10722599_7

[27] Sandhu, R. S. The schematic protection model: its definition and analysis for acyclic attenuating schemes. Journal of the ACM (JACM), 35(2):404–432, 1988. https://doi.org/10.1145/42282.42286

[28] Ferraiolo, D. et Kuhn, R. Role-based access control. In In 15th NIST-NCSC National Computer Security Conference, (1992). https://doi.org/10.1016/S0065-2458(08)60206-5

[29] Loscocco, P. A., Smalley, S. D., Muckelbauer, P. A., Taylor, R. C., Turner, S. J. et Farrell, J. F. The inevitability of failure: The flawed assumption of security in modern computing environments. In Proceedings of the 21st National Information Systems Security Conference, volume 10, pages 303–314, 1998.

[30] Anderson, J. P. Computer security threat monitoring and surveillance. Rapport technique, Technical report, James P. Anderson Company, Fort Washington, Pennsylvania, 1980.

[31] Bell, D. E. et LaPadula, L. J. Secure computer systems: Mathematical foundations. Rapport technique, DTIC Document, 1973.

[32] Biba, K. J. Integrity considerations for secure computer systems. Rapport technique, DTIC Document, 1977.

[33] Boebert, W. E. et Kain, R. Y. A practical alternative to hierarchical integrity policies. NIST SPECIAL PUBLICATION SP, pages A–10, 1989.

[34] Sandhu, R. S., Coyne, E. J., Feinstein, H. L. et Youman, C. E. Role-based access control models. Computer, (2):38–47, 1996. https://doi.org/10.1109/2.485845

[35] Sandhu, R., Ferraiolo, D. et Kuhn, R. The nist model for role-based access control: towards a unified standard. In ACM workshop on Role-based access control, volume 2000. https://doi.org/10.1145/344287.344301

[36] Standard, R. Incits 359-2004. ANSI INCITS, pages 359–2004, 2004.

[37] Kalam, A. A. E., Baida, R., Balbiani, P., Benferhat, S., Cuppens, F., Deswarte, Y., Miege, A., Saurel, C. et Trouessin, G. Organization based access control. In Policies for Distributed Systems and Networks, 2003. Proceedings. POLICY 2003. IEEE 4th International Workshop on, pages 120–131. IEEE, 2003. https://doi.org/10.1109/POLICY.2003.1206966

[38] Hu, V. C., Ferraiolo, D., Kuhn, R., Schnitzer, A., Sandlin, K., Miller, R. et Scarfone, K. Guide to attribute based access control (abac) definition and considerations. NIST Special Publication, 800:162, 2014. https://doi.org/10.6028/NIST.SP.800-162

[39] Jin, X., Krishnan, R. et Sandhu, R. S. A unified attributebased access control model covering dac, mac and rbac. DBSec, 12:41–55, 2012. https://doi.org/10.1007/978-3-642-31540-4_4

[40] Moses, T. et al. Extensible access control markup language (xacml) version 2.0. Oasis Standard, 200502, 2005.

[41] OASIS. OASIS. https://www.oasis-open.org/. 2015.

[42] Hu, V. C., Martin, E., Hwang, J. et Xie, T. Conformance checking of access control policies specified in xacml. In Computer Software and Applications Conference, 2007. COMPSAC 2007. 31st Annual International, volume 2, pages 275–280. IEEE, 2007. https://doi.org/10.1109/COMPSAC.2007.96
[43] Damianou, N., Dulay, N., Lupu, E. et Sloman, M. A language for specifying security and management policies for distributed systems. London: Department of Computing, Imperial College, Tech. Rep, 2000.
[44] Twidle, K., Dulay, N., Lupu, E. et Sloman, M. Ponder2: A policy system for autonomous pervasive environments. In Autonomic and Autonomous Systems, 2009. ICAS'09. Fifth International Conference on, pages 330–335. IEEE, 2009. https://doi.org/10.1109/ICAS.2009.42
[45] Waltermire, D., Schmidt, C., Scarfone, K. et Ziring, N. Specification for the extensible configuration checklist description format (xccdf) version 1.2 (draft). (2011b).
[46] Waltermire, D., Quinn, S., Scarfone, K. et Halbardier, A. The technical specification for the Security Content Automation Protocol (SCAP). NIST Special Publication, 800:126. (2011a).
[47] Barrère, M., Badonnel, R. et Festor, O. Collaborative remediation of configuration vulnerabilities in autonomic networks and systems. In Proceedings of the 8th International Conference on Network and Service Management, pages 357–363. International Federation for Information Processing. (2012a).
[48] OVAL Oval language. Http://oval.mitre.org/. 2014.
[49] Barrère, M., Badonnel, R. et Festor, O. Towards the assessment of distributed vulnerabilities in autonomic networks and systems. In Network Operations and Management Symposium (NOMS), 2012 IEEE, pages 335–342. IEEE. (2012b). https://doi.org/10.1109/NOMS.2012.6211916
[50] Azraoui, M., Elkhiyaoui, K., Önen, M., Bernsmed, K., De Oliveira, A. S. et Sendor, J. A-ppl : An accountability policy language. In Data Privacy Management, Autonomous Spontaneous Security, and Security Assurance, pages 319–326. Springer, 2015. https://doi.org/10.1007/978-3-319-17016-9
[51] A., Erraissi A., Belangour A. Capturing Hadoop Storage Big Data Layer Meta-Concepts. In: Ezziyyani M. (eds) Advanced Intelligent Systems for Sustainable Development (AI2SD'2018). AI2SD 2018. Advances in Intelligent Systems and Computing, vol 915. Springer, Cham, 2019. https://doi.org/10.1007/978-3-030-11928-7_37
[52] A. Erraissi and A. Belangour, "Meta-modeling of Zookeeper and MapReduce processing," 2018 International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), Kenitra, Morocco, pp. 1-5, 2018. https://doi.org/10.1109/ICECOCS.2018.8610630
[53] Rodriguez, A., Fern´andez-Medina, E., and Piattini, M. A bpmn extension for the modeling of security requirements in business processes. IEICE Transactions, 90-D(4):745–752, 2007. https://doi.org/10.1093/ietisy/e90-d.4.745
[54] Mellado, D., Fernandez-Medina, E., and Piattini, M. A common criteria based security requirements engineering process for the development of secure information systems. Comput. Stand. Interfaces, 29(2):244–253, 2007. https://doi.org/10.1016/j.csi.2006.04.002
[55] Samarati, P. and Capitani, S. D. Access control: Policies, models, and mechanisms. In FOSAD, pages 137–196, 2000. https://doi.org/10.1007/3-540-45608-2_3
[56] Ferraiolo, D. et Kuhn, R. Role-based access control. In In 15th NIST-NCSC National Computer Security Conference, 1992.
[57] Loscocco, P. A., Smalley, S. D., Muckelbauer, P. A., Taylor, R. C., Turner, S. J. et Farrell, J. F. (1998). The inevitability of failure: The flawed assumption of security in modern computing environments. In Proceedings of the 21st National Information Systems Security Conference, volume 10, pages 303–314.
[58] A. Erraissi, B. Mouad and A. Belangour, "A Big Data visualization layer meta-model proposition," 2019 8th International Conference on Modeling Simulation and Applied Optimization (ICMSAO), Manama, Bahrain, 2019, pp. 1-5. doi: 10.1109/ICMSAO.2019.8880276
[59] M. Banane, A. Erraissi and A. Belangour, "SPARQL2Hive: An approach to processing SPARQL queries on Hive based on meta-models," 2019 8th International Conference on Modeling Simulation and Applied Optimization (ICMSAO), Manama, Bahrain, 2019, pp. 1-5. doi: 10.1109/ICMSAO.2019.8880393
[60] Fatima Kalna, Allae Erraissi, Mouad Banane, Belangour "A Scalable Business Intelligence Decision-Making System in The Era of Big Data" International Journal of Innovative Technology and Exploring Engineering 2019. https://doi.org/10.35940/ijitee.L3251.1081219