



ASTES

# Advances in Science, Technology & Engineering Systems Journal

---

VOLUME 8-ISSUE 1 | JAN-FEB 2023

[www.astesj.com](http://www.astesj.com)

ISSN: 2415-6698

## **EDITORIAL BOARD**

### **Editor-in-Chief**

**Prof. Passerini Kazmerski**  
University of Chicago, USA

### **Editorial Board Members**

**Dr. Serdar Sean Kalaycioglu**  
Toronto Metropolitan University, Canada

**Dr. Heba Afify**  
MTI university, Cairo, Egypt

**Mr. Randhir Kumar**  
National University of Technology Raipur, India

**Dr. Jiantao Shi**  
Nanjing Research Institute of Electronic Technology, China

**Dr. Tariq Kamal**  
University of Nottingham, UK  
Sakarya University, Turkey

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

**Dr. Nguyen Tung Linh**  
Electric Power University, Vietnam

**Prof. Majida Ali Abed Meshari**  
Tikrit University Campus, Iraq

**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

**Mr. Muhammad Tanveer Riaz**  
School of Electrical Engineering, Chongqing University, P.R. China

**Dr. Daniele Mestriner**  
University of Genoa, Italy

**Dr. Hung-Wei Wu**  
Kun Shan University, Taiwan

### **Regional Editors**

**Dr. Ahmet Kayabasi**  
Karamanoglu Mehmetbey University, Turkey

**Dr. Maryam Asghari**  
Shahid Ashrafi Esfahani, Iran

**Dr. Shakir Ali**  
Aligarh Muslim University, India

**Mr. Aamir Nawaz**  
Gomal University, Pakistan

**Dr. Ebubekir Altuntas**  
Gaziosmanpasa University, Turkey

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

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

**Dr. Gomathi Periasamy**  
Mekelle University, Ethiopia

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

**Mr. Manu Mitra**  
University of Bridgeport, USA

**Mr. Abdullah El-Bayoumi**  
Cairo University, Egypt

**Dr. Ayham Hassan Abazid** Jordan  
University of Science and Technology, Jordan

**Dr. Qichun Zhang**  
University of Bradford, United Kingdom

## Editorial

In the ever-evolving landscape of scientific inquiry, this issue brings together a diverse collection of research papers spanning various domains of technology and science. Each paper contributes valuable insights and innovative approaches to address contemporary challenges. This editorial provides a brief overview of 16 accepted papers, highlighting its key contributions and implications.

The first paper explores the escalating energy consumption in cloud data centers, a consequence of the increasing number of servers. Tackling the problem through meta-heuristic and heuristic algorithms, the authors propose an approach using Genetic Algorithm (GA) and First Fit Decreasing (FFD) for workload placement and power consumption prediction in data centers [1]. The study demonstrates the efficiency of the proposed algorithms, particularly the superiority of GA compared to Ant Colony Optimization (ACO) and Simulated Annealing (SA).

Moving to the realm of nanotechnology, the second paper delves into the characterization of Co-Axial Cylindrical Carbon Nanotube Field Effect Transistor (CNTFET). Highlighting the advantages of CNTFET over traditional Si-MOSFET, the research investigates the impact of gate-insulator thickness on its performance, providing valuable insights for optimizing its current carrying capacity [2].

Shifting gears to robotics, the third paper introduces an open-source anthropomorphic robotic hand designed for telepresence robots. Utilizing a four-bar linkage mechanism to reduce the number of actuators while maintaining human-like movement, the authors present a cost-effective solution with applications in various fields [3]. The experimental results demonstrate the hand's capability to perform human-like movements and grasp various objects.

The fourth paper addresses the critical issue of droughts, focusing on the Marathwada region. Through a comprehensive analysis of rainfall data spanning nearly four decades, the study identifies negative trends in annual rainfall, particularly in the monsoon season. The findings lay the groundwork for advanced computation techniques to predict and mitigate droughts [4].

Shifting to the domain of error-correcting codes, the fifth paper presents a new decoder for Reed Solomon and BCH codes using a novel syndromes block. The proposed algorithm aims to reduce the number of iterations compared to existing blocks, demonstrating its efficacy through hardware description language implementation and simulation [5].

The sixth paper explores the realm of robotics control algorithms, introducing Optimal Control Allocation (OCA) and Nonlinear Model Predictive Control (NMPC) for a rover robotics system with mecanum wheels and dual arms. The study showcases the efficiency of these algorithms in addressing wheel and joint torque saturation issues while manipulating a heavy payload [6].

The seventh paper introduces an innovative approach to edge detection-based image steganography. The authors propose hybridizing edge detectors dynamically based on embedding demands, using logical AND, OR, or OR with dilation operations. The results demonstrate improved embedding capacity and security against attacks compared to existing methods [7].

In the eighth paper, the focus shifts to the robust  $H^\infty$  control of nonlinear systems subject to cyber-attacks. The authors present a polytopic modeling approach and a robust controller design

method, demonstrating its efficacy in ensuring stability and security for a quadrotor/UAV subject to actuator attacks [8].

The ninth paper introduces an ensemble of voting-based deep learning models with regularization functions for sleep stage classification. Leveraging recurrent neural network (RNN), long short-term memory (LSTM), and gated recurrent unit (GRU) models, the study achieves impressive accuracy in sleep stage classification [9].

Moving to the field of subsurface utility engineering (SUE), the tenth paper proposes an integrated GIS-SUE map cost estimation system prototype. By utilizing GIS-compatible digital spatial maps, the research aims to enhance the efficiency of utility mapping and cost estimation, providing a valuable tool for managing and maintaining utilities [10].

The eleventh paper explores the realm of olfactory acquisition through the conception and simulation of an electronic nose prototype. The study focuses on designing an efficient gas chamber for gas sensors, enhancing the overall performance of the electronic nose in measuring gas quality [11].

The twelfth paper addresses the pressing need for strengthening cybersecurity management in public organizations. The research develops a model to identify the cybersecurity management capacity of public organizations, providing a process for assessment and categorization based on the level of cybersecurity capacity [12].

The thirteenth paper contributes to the field of power system optimization, comparing the performance of particle swarm optimization (PSO) and firefly algorithm (FA) in the optimal allocation of a static synchronous compensator (STATCOM). The study showcases the benefits of FA over PSO in reducing voltage deviation and power losses [13].

In the realm of agricultural technology, the fourteenth paper introduces a novel deep learning method for the detection of Northern Leaf Blight and Gray Leaf Spot in corn. Leveraging YOLOv3 with additional enhancements, the research presents a valuable tool for early disease detection in corn crops [14].

The fifteenth paper introduces an active simulation of grounded parallel-type immittance functions employing voltage differencing buffered amplifiers (VDBAs) and all grounded passive components. The proposed circuit demonstrates the feasibility of simulating parallel-type impedances with only two VDBAs and two grounded passive components [15].

Finally, the sixteenth paper addresses the urgent need for effective models in teaching mathematics to gifted students. The research develops a comprehensive model based on various teaching approaches, demonstrating its effectiveness in improving the performance of gifted students [16].

In conclusion, this compilation of research papers reflects the richness and diversity of contemporary scientific endeavors. Each contribution brings forth innovative solutions, insights, and methodologies, contributing to the collective pursuit of knowledge and advancement in their respective fields.

## References:

- [1] A. Bouaouda, K. Afdel, R. Abounacer, "Meta-heuristic and Heuristic Algorithms for Forecasting Workload Placement and Energy Consumption in Cloud Data Centers," *Advances in Science, Technology and Engineering Systems Journal*, **8**(1), 1–11, 2023, doi:10.25046/aj080101.
- [2] S. Sen, A. Sarkar, P. Chakraborty, "Characterization and Investigating the Effect of Gate-Insulator Thickness on Co-Axial Cylindrical Carbon Nanotube Field Effect Transistor," *Advances in Science, Technology and Engineering Systems Journal*, **8**(1), 12–16, 2023, doi:10.25046/aj080102.
- [3] J. Trichada, T. Wimonrut, N. Tirasuntarakul, E. Pengwang, "Design of an Open Source Anthropomorphic Robotic Hand for Telepresence Robot," *Advances in Science, Technology and Engineering Systems Journal*, **8**(1), 17–26, 2023, doi:10.25046/aj080103.
- [4] H. Bana, R.D. Garg, "Analysis and Trend Estimation of Rainfall and Seasonality Index for Marathwada Region," *Advances in Science, Technology and Engineering Systems Journal*, **8**(1), 30–37, 2023, doi:10.25046/aj080104.
- [5] M. Elghayaty, A.E.H. El Idrissi, O. Mouhib, A. Wahbi, A. Hadjoudja, "Design, Optimization and Simulation of a New Decoder for Reed Solomon and BCH Codes using the New Syndromes Block," *Advances in Science, Technology and Engineering Systems Journal*, **8**(1), 38–43, 2023, doi:10.25046/aj080105.
- [6] S. Kalaycioglu, A. de Ruiter, "Nonlinear Model Predictive Control of Rover Robotics System," *Advances in Science, Technology and Engineering Systems Journal*, **8**(1), 44–56, 2023, doi:10.25046/aj080106.
- [7] H. Sultana, A.H.M. Kamal, "An Efficient Way of Hybridizing Edge Detectors Depending on Embedding Demand," *Advances in Science, Technology and Engineering Systems Journal*, **8**(1), 63–77, 2023, doi:10.25046/aj080108.
- [8] B.-R. Souad, "On the Polytopic Modelling & Robust  $H^\infty$  Control of Nonlinear Systems Subject to Cyber-attack: Application to Attitude Stabilization of Quadrotor," *Advances in Science, Technology and Engineering Systems Journal*, **8**(1), 78–83, 2023, doi:10.25046/aj080109.
- [9] S. Kaliyapillai, S. Krishnamurthy, T. Murugasamy, "An Ensemble of Voting- based Deep Learning Models with Regularization Functions for Sleep Stage Classification," *Advances in Science, Technology and Engineering Systems Journal*, **8**(1), 84–94, 2023, doi:10.25046/aj080110.
- [10] A. Nashwan, K. Al-Joburi, "Integrated GIS-SUE Map Cost Estimation System Prototype for Designing a Decision Support System," *Advances in Science, Technology and Engineering Systems Journal*, **8**(1), 95–100, 2023, doi:10.25046/aj080111.
- [11] M. Harmouzi, A. Amari, L. Masmoudi, "Conception and Simulation of an Electronic Nose Prototype for Olfactory Acquisition," *Advances in Science, Technology and Engineering Systems Journal*, **8**(1), 101–107, 2023, doi:10.25046/aj080112.
- [12] R.R. Izurieta, S.M.T. Toapanta, L.J.C. Morales, M.M.B. Hifóng, E.Z.G. Díaz, O.M.Z. Vizuete, L.E.M. Gallegos, J.A.O. Trejo, "Prototype to Identify the Capacity in Cybersecurity Management for a Public Organization," *Advances in Science, Technology and Engineering Systems Journal*, **8**(1), 108–115, 2023, doi:10.25046/aj080113.
- [13] A. Jimoh, S.O. Ayanlade, E.I. Ogunwole, D.E. Owolabi, A.B. Jimoh, F.M. Aremu, "Metaheuristic Optimization Algorithm Performance Comparison for Optimal Allocation of Static Synchronous Compensator," *Advances in Science, Technology and Engineering Systems Journal*, **8**(1), 116–124, 2023, doi:10.25046/aj080114.
- [14] B. Song, J. Lee, "Northern Leaf Blight and Gray Leaf Spot Detection using Optimized YOLOv3," *Advances in Science, Technology and Engineering Systems Journal*, **8**(1), 125–130, 2023, doi:10.25046/aj080115.
- [15] P. Mongkolwai, P. Moonmuang, W. Tangsrirat, T. Suesut, "Active Simulation of Grounded Parallel-Type Immittance Functions Employing VDBAs and All Grounded Passive Components," *Advances in Science, Technology and Engineering Systems Journal*, **8**(1), 131–137, 2023, doi:10.25046/aj080116.

- [16] Z. Dedovets, M. Rodionov, A. Novichkova, "A Model for Teaching Mathematics to Gifted Students Based on an Effective Combination of Various Approaches for their Preparation," *Advances in Science, Technology and Engineering Systems Journal*, **8**(1), 138–148, 2023, doi:10.25046/aj080117.

**Editor-in-chief**

**Prof. Passerini Kazmersk**

# ADVANCES IN SCIENCE, TECHNOLOGY AND ENGINEERING SYSTEMS JOURNAL

Volume 8 Issue 1

January-February 2023

## CONTENTS

<i>Meta-heuristic and Heuristic Algorithms for Forecasting Workload Placement and Energy Consumption in Cloud Data Centers</i> Amine Bouaouda, Karim Afdel, Rachida Abounacer	01
<i>Characterization and Investigating the Effect of Gate-Insulator Thickness on Co-Axial Cylindrical Carbon Nanotube Field Effect Transistor</i> Suchismita Sen, Argha Sarkar, Pinaki Chakraborty	12
<i>Design of an Open Source Anthropomorphic Robotic Hand for Telepresence Robot</i> Jittaboon Trichada, Traithep Wimonrut, Narongsak Tirasuntarakul, Eakkachai Pengwang	17
<i>Analysis and Trend Estimation of Rainfall and Seasonality Index for Marathwada Region</i> Himanshu Bana, Rahul Dev Garg	30
<i>Design, Optimization and Simulation of a New Decoder for Reed Solomon and BCH Codes using the New Syndromes Block</i> Mohamed Elghayyaty, Anas El Habti El Idrissi, Omar Mouhib, Azeddine Wahbi, Abdelkader Hadjoudja	38
<i>Nonlinear Model Predictive Control of Rover Robotics System</i> Serdar Kalaycioglu, Anton de Ruyter	44
<i>Economic-ecological study of the solutions proposed by HOMER for the optimization of an industrial installation</i> Othmane Echarradi, Abdessamad Benlafkih, Mounir Fahoume	Withdrawn
<i>An Efficient Way of Hybridizing Edge Detectors Depending on Embedding Demand</i> Habiba Sultana, A. H. M. Kamal	63
<i>On the Polytopic Modelling &amp; Robust <math>H^\infty</math> Control of Nonlinear Systems Subject to Cyber-attack: Application to Attitude Stabilization of Quadrotor</i> Bezzaoucha-Rebaï Souad	78
<i>An Ensemble of Voting- based Deep Learning Models with Regularization Functions for Sleep Stage Classification</i> Sathyabama Kaliyapillai, Saruladha Krishnamurthy, Thiagarajan Murugasamy	84
<i>Integrated GIS-SUE Map Cost Estimation System Prototype for Designing a Decision Support System</i> Ali Nashwan, Khalil Al-Joburi	95

<i>Conception and Simulation of an Electronic Nose Prototype for Olfactory Acquisition</i>	101
Mostapha Harmouzi, Aziz Amari, Lhoussaine Masmoudi	
<i>Prototype to Identify the Capacity in Cybersecurity Management for a Public Organization</i>	108
Richard Romero Izurieta, Segundo Moisés Toapanta Toapanta, Luis Jhony Caucha Morales, María Mercedes Baño Hifóng, Eriannys Zharayth Gómez Díaz, Oscar Marcelo Zambrano Vizúete, Luis Enrique Mafla Gallegos, José Antonio Orizaga Trejo	
<i>Metaheuristic Optimization Algorithm Performance Comparison for Optimal Allocation of Static Synchronous Compensator</i>	116
Abdulrasaq Jimoh, Samson Oladayo Ayanlade, Emmanuel Idowu Ogunwole, Dolapo Eniola Owolabi, Abdulsamad Bolakale Jimoh, Fatina Mosunmola Aremu	
<i>Northern Leaf Blight and Gray Leaf Spot Detection using Optimized YOLOv3</i>	125
Brian Song, Jeongkyu Lee	
<i>Active Simulation of Grounded Parallel-Type Immittance Functions Employing VDBAs and All Grounded Passive Components</i>	131
Pratya Mongkolwai, Pitchayanin Moonmuang, Worapong Tangsirat, Taweepol Suesut	
<i>A Model for Teaching Mathematics to Gifted Students Based on an Effective Combination of Various Approaches for their Preparation</i>	138
Zhanna Dedovets, Mikhail Rodionov, Anna Novichkova	



# Meta-heuristic and Heuristic Algorithms for Forecasting Workload Placement and Energy Consumption in Cloud Data Centers

Amine Bouaouda\*, Karim Afdel, Rachida Abounacer

<sup>1</sup>Department of Computer Science, LabSIV, Ibn Zohr University, Agadir, 80000, Morocco

## ARTICLE INFO

Article history:

Received: 30 August, 2022

Accepted: 03 December, 2022

Online: 24 January, 2023

Keywords:

Data center Energy

Forecasting Energy

Container Placement

Genetic Algorithm

First Fit Decreasing

## ABSTRACT

The increase of servers in data centers has become a significant problem in recent years that leads to a rise in energy consumption. The problem of high energy consumed by data centers is always related to the active hardware especially the servers that use virtualization to create a cloud workspace for the users. For this reason, workload placement such as virtual machines or containers in servers is an essential operation that requires the adoption of techniques that offer practical and best solutions for the workload placement that guarantees an optimization in the use of material resources and energy consumption in the cloud. In this article, we propose an approach that uses heuristics and meta-heuristics to predict cloud container placement and power consumption in data centers using a Genetic Algorithm (GA) and First Fit Decreasing (FFD). Our algorithms have been tested on CloudSim and the results showed that our methods gave better and more efficient solutions, especially the Genetic Algorithm after comparing them with Ant Colony Optimization (ACO) and Simulated Annealing (SA).

## 1 Introduction

This paper is an extension of work originally presented in the Fifth Conference on Cloud and Internet of Things (CIoT) [1]. By adding other methods such as the Genetic Algorithm, First-Fit, Random-Fit, and Simulated Annealing, our approach will predict the energy consumed by the data centers and the workload placement (containers) in the servers, offering best and optimal solutions to reduce energy consumption, waste of cloud resources, and have an energy-efficient container placement policy.

Generally, Cloud Computing is based on a large data center (server farm), in which many servers are connected to achieve high performance. The data centers represent an infrastructure of several instances (hosts, virtual machines...) [2]–[4]. Each of these instances requires an allocation at the data center because of the growing demand for hosting services. For these reasons, they are considered to be heavy consumers of resources and energy [3]. In data center, the energy consumed by active servers represents a large proportion of the total energy [4], [5]. More clearly, the energy consumed by the hosts or hosting servers plus network and storage equipment represents about 40% of the total energy [6]–[8]. Cooling equipment uses between 45% and 50% of the total energy, and the rest is shared among other systems such as lighting [6]–[11].

With the energy of the cooling systems, the costs in the data centers are experiencing a big explosion, which require a reduction in their expenses [12]. According to [5][9][13], the main challenges in data centers are to minimize the heat and energy consumed by cloud infrastructures and to secure them against threats.

So, to optimize the use of energy in data centers, it is necessary to define the servers that must be active according to the current workload and to avoid traditional techniques that negatively influence the quality of services (QoS) such as stopping components or reducing their performance [6], [9], [10]. In most data centers, the consumption of hardware and software resources of each active physical machine is between 11% and 50% with power consumption between 50% and 70% compared to a server whose resources are used entirely by the hosting of the instances and applications executed on these instances [2], [9], [14]. For this, the efficient placement of virtual instances is very important to control the use of material resources and prohibit any kind of their waste that can lead to an increase in energy consumption.

Our approach will have two objectives. The first will be the prediction of workload placement using the Genetic Algorithm and the First Fit Decreasing to define a better placement of a new type of virtual instance called containers. This placement will be constrained to define the best lower number of servers to host container

\*Corresponding Author: Amine Bouaouda, +212604129135, amine.bouaouda@edu.uiz.ac.ma

instances. The second objective is the main one, which will be the prediction of energy consumed by a data center, by applying our container placement algorithms to have the best solutions in terms of reducing the energy consumed by a data center and optimizing the use of hardware resources such as RAM, Storage, Bandwidth, etc.

The remainder of this paper is coordinated as follows. We present the related works in the second section. In the third section, we propose our methodology. We perform our algorithms in the fourth section, and the paper is concluded in section five.

## 2 Related Work

The estimation or prediction of energy consumption in data centers has become a necessary operation in recent years. Due to the huge growth in user demands for cloud services, the number of servers has started to increase to provide the necessary hardware resources. This implied an increase in energy consumption, which forced large cloud companies to do studies on the energy consumed by their data centers, to optimize it in the future, or replace their power source with a cheaper and guaranteed one.

In this context, several approaches estimate energy consumption in data centers. Most of these methods focus on servers to estimate energy. Each server or host is a set of resources such as RAM, HDD, and CPU. In this case, the energy of the host is relative to the sum of the power of all its resources, or according to some [15][16]. Mathematically, the energy consumed by a host is represented in the literature by linear functions that depend on a resource like CPU [15][17], or non-linear, whose functions are quadratic of the CPU resource use [17][18]. Heuristics and metaheuristics are methods that are widely used to estimate the energy in data centers based on the placement of virtual instances in servers especially metaheuristics as they are adaptive for complex problems that require considerable calculation.

In [19], they proposed the placement of virtual machines based on energy consumption by the resources of the data center. One of the objectives of this approach was to reduce energy consumption using a simulated annealing algorithm. This technique generates an initial solution called initial configuration which contains the placement of the VMs in servers, on which they applied at each iteration one of the three simulated annealing techniques: inversion, translation, and switching, to get the next configuration. To define which solution is better, they calculate the energy consumed by the data center in both cases and choose the one that gives the small energy value that will be the best solution. The disadvantages of this approach are the execution time which is very long for large instances, and its process ends if it finds that a new configuration is better than the previous one even if they remain several iterations. This decision does not necessarily indicate that the new configuration is the best among all other solutions.

The same thing in [20], they used the simulated annealing algorithm to propose an economic placement in terms of energy for virtual machines. The proposed algorithm goes through the four stages of the simulated annealing (generation of the initial configuration, obtaining the next generation, definition of the objective function, and timing of temperatures and evolution time). The simu-

lated annealing proposed was compared by the First Fit Decreasing multi-start random searching approach. The results obtained show the effectiveness of the proposed method, but in the case of large instances, it takes a lot of time to find the best solution. For this, they used a time limiter to stop the calculation process in a solution very close to the best.

Other researchers have proposed multi-objective metaheuristics for the placement of workloads in servers by calculating the energy consumed by data centers, to select the best placement in terms of energy minimization. In [21], the authors used the multi-objective genetic algorithm to provide virtual machine placement by minimizing energy consumption and improving the quality of services and the use of resources. The contradictory nature of the objectives defined in this approach has necessarily influenced the distribution of the load between the resources in a data center.

The same thing in [22] and [23], which proposed an approach based on the ant colony algorithm to calculate and minimize the energy consumed by a cloud system based on the placement of virtual machines. In [23], they built a mechanism for measuring indirect energy consumption for virtual machines based on a model for calculating the energy consumed by these machines deployed in [24], because it is difficult to deduce the energy consumption directly from the material because of the existence of the virtualization technology.

Our approach will be different from the old works because we will use cloud containers instead of virtual machines. More of this, we will propose a genetic algorithm and a First-Fit-Decreasing algorithm to provide best container placement without wasting the resources of active servers, with the prediction of energy consumed by any data center to define which algorithm offers best and optimal solutions for the minimization of energy consumed and the optimization of material resources.

## 3 Methodology

Our main objective is to predict the energy consumed in a data center for a given workload using the genetic algorithm and First-Fit-Decreasing to predict the best or optimal placement of this load (cloud containers). The best algorithm will be one that offers optimal solutions in terms of minimizing energy consumption and optimizing the use of the hardware resources of a data center operating in a specific context.

### 3.1 Workload Placement

The placement of workloads is an operation that has a great impact on several problems in cloud computing such as minimizing energy consumption. The concept of this placement is to define the servers that will be active to host several virtual instances. To have an optimal placement, it is necessary to choose an optimal number of servers without wasting the material resources of the data center. The workloads in our approach will be the containers.

Containers are small virtual instances in terms of hardware and software resources [25]–[27]. They provide a virtual platform such as Docker, with which multiple users can drive and run their applications or images of operating systems directly on the physical

machine [27], [26].

The containers are efficient compared to the virtual machines because they are lightweight and their installation takes a few seconds and they run directly on the operating system and the hardware of the physical machine. On the other hand, virtual machines take a long time to install and require hypervisors to run [26][25]. In [26], the author represents a containerization technology for creating container instances. It allows users to deploy and run their applications in process containers.

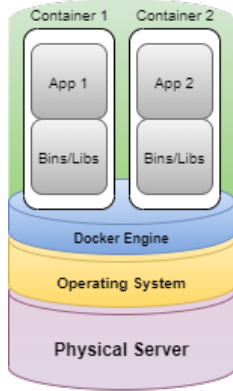


Figure 1: Operation of a container

Our approach will propose two algorithms to predict the best and optimal placement for containers. The main objective will be the choice of an optimal number of servers sufficient to host the containers according to their material resources. This objective is defined by the following equation:

$$F(X) = \sum_{Host=1}^n X_{Host} \quad (1)$$

$$X_{Host} = \begin{cases} 1, & \text{if the host is active} \\ 0, & \text{otherwise} \end{cases}$$

To achieve this goal, we have identified three necessary constraints. First, each container must have a placement at the end. The second constraint insists that each container must be placed in a single host. The latter checks that all containers must not exceed the host regarding material resources. For example, for RAM, we have decided to reserve 80% of each active host for container placement and keep the rest for user processing, ensuring proper load balancing between servers.

### 3.1.1 Genetic Algorithm

The genetic algorithm [28] is an optimization method (meta-heuristic) [29] first presented by John Holland in the 1970s. It is based on techniques derived from genetics and the evolutionary mechanisms of nature: crossover, mutation, and selection.

It represents a method for solving complex optimization problems, with or without constraints, based on a natural selection process (a process similar to that of biological evolution). More precisely,

it provides solutions to problems that don't have calculable solutions in a reasonable time analytically[30][28].

The process of the genetic algorithm begins with the random creation of thousands of more or less good solutions, then they will be subjected to an evaluation to select the most suitable ones according to constraints. The population continues to evolve through the creation of other generations, by crossing the best solutions between them and having them mutate, then they are brought together with the best already chosen in the selection. This process will be restarted in a certain number of iterations to arrive at the best solutions.

**Generation of the initial population:** The genetic algorithm in its nature is a population-based method, i.e., it begins with a set of initial solutions named the initial population, and in each iteration, it produces a new generation of solutions of the same size as the initial one. To have an initial population, it must be generated from a set of solutions, which are called individuals (chromosomes). The total number of individuals generated represents the size of the population.

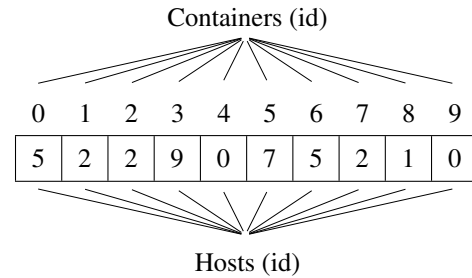


Figure 2: Representation of a chromosome (individual)

So to generate the initial population, you have to create a set of chromosomes as in the figure 3. This generation will be random within the constraints of the approach to ensure that each respects all the constraints of the problem to have a correct initial population.

	0	1	2	3	4	5	6	7	8	9
Chromosome 0	5	2	2	9	0	7	5	2	1	0
Chromosome 1	9	5	0	1	6	2	5	2	1	8
Chromosome 2	2	2	4	7	3	7	5	2	1	0
Chromosome 3	8	0	5	4	4	3	0	2	1	0

Figure 3: Representation of the structure of an initial population

**Selection:** For our approach, we decided to set the number of chromosomes in four for the initial population and the one to be built in each iteration and select the two best individuals (parents) among each created population, on which the crossing operator will apply to have two new individuals (sons) to build a new generation with new features. The selection of the best individuals is based on a criterion called the Fitness function. It is an equation defined

according to the problem studied, to determine the best solutions that satisfy this function.

To determine the best individuals, we have based on the principle of our objective which serves to minimize the number of active hosts. More clearly, if a chromosome of a population respects all the constraints of the problem and contains the minimum possible hosts, then it represents a good solution among those in the studied population. This choice makes it possible to calculate the number of active hosts in an individual and choose two individuals which contain the minimum number of hosts, or the reciprocal of the number of active hosts, and then choose the two large values obtained. We represent below the equation of the Fitness function to select the best individuals.

$$f(I) = 1/N_{AH} \quad (2)$$

- $I$  : The individual being studied.
- $f(I)$  : The Fitness function of individual  $I$ .
- $N_{AH}$  : Number of active hosts in the individual  $I$ .

The Fitness function will be applied at the beginning of the initial population to determine the individual parents, who will be the inputs of the next phase (the crossover). This function will be applied to each new generation created until the last iteration. These kinds of individuals, who represent the parents are feasible (workable) solutions for the problem studied. Workable individuals are solutions that met the needs of a problem.

So, the individual (parents) selected according to the Fitness function in each iteration are feasible solutions among others to discover probably in the following iterations. But instead of choosing both parents, we decided to compare them by choosing the best among them (the one that has the greatest value of Fitness, if they have equal values, then both will be chosen as feasible solutions). And each feasible solution represents one of the best-suggested ways to place containers in a minimum number of hotels without any waste of resources or energy consumed.

**Crossover:** The choice of the best individuals in the selection phase is the starting point of the crossover. These selected individuals are genetically better according to the function of Fitness, and they contain characteristics that will improve each population by producing new individuals called sons. The creation of new individuals is done by a crossover applied to the parents explained by the following two points:

- Divide each parent individual into two parts (from the middle), if the size of the individual is even, then both parts will have the same length, otherwise, the first part will exceed the size of the second with an element.
- Concatenate the first part of the first parent with the second of the other to have the first son, then the second part of the first parent with the first of the other parent to have the second son.

After creating the new individuals (sons), the parent individuals are kept to build a new generation of four chromosomes as the initial

population after the application of the mutation operation on the sons. Keeping parents is not optional, but it has two very important roles in the algorithm process.

First, it avoids the case of having a bad generation, because parents are already good solutions according to the Fitness function applied in the selection, on the other hand, there is no information about the nature of new individuals created, whether they are good solutions to the problem or not, until the application of the mutation phase and build a new population and apply the selection operator to it to indicate the individuals who will be the new parents who may be one of the created sons of the previous iteration or both or none.

Secondly, the conservation of parents and combining them with the new individuals makes it possible to respect the rule of construction of a population in the genetic algorithm which indicates that all generations must be of the same size as the one at the beginning, as well as this combination guarantees diversity in the best solutions and increase the proportion of having several that are feasible.

**Mutation:** Generally, the mutation principle is used to apply more change to the sons created in the crossover phase to obtain a rich and different population from the previous one. These changes will apply to some genes of one or more chromosomes. In our case, the genes on each chromosome are the identifiers of the hosts chosen to host the containers, as shown in the figure 4.

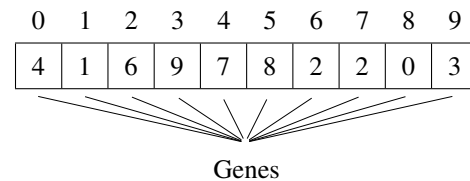


Figure 4: Individual-level active hosts represent their genes

So the change in the hosts of an individual involves a modification in the location of the containers. For this reason, the mutation phase for our approach will be slightly different compared to those of other problems that use the genetic algorithm. More precisely, The main role of the mutation will be corrective, that is to correct the faults produced by the crossing at the level of the son individuals. This crossing can cause false locations for containers at these individuals. For this reason, we must first check the location of the containers in each child's individual if one or more exceed the host that hosts them in terms of RAM. In this way, the two sons will be corrected by ensuring that they respect the constraints of the approach before combining them with the parents to build a new generation.

This "corrective" mutation has another impact because the correction made at the level of the sons will have a great chance of bringing back new characteristics. After all, the containers in those chromosomes have probably different locations to those in the individual parents, and therefore the possibility of having other better solutions. With all these characteristics, the mutation phase will have a great impact on the realization of a good dynamic container placement which is one of the big goals of this optimization problem, meaning the appearance of the migration notion of virtual instances

between active hosts, which is absent in the static placement which prevents the migration of a container from one host to another after having placed it.

---

**Algorithm 1:** Container Placement based on Genetic Algorithm
 

---

**Data:** List of container requests and list of hosts  
**Result:** Deploy containers on hosts  
 $HL \leftarrow HostsList;$   
 $CL \leftarrow ContainersList;$   
**for**  $i = 1$  **to** 4 **do**  
   $list_{InitialSolutions} \leftarrow createChromosome(HL, CL);$   
   $map_{InitialPopulation}.add(i, list_{InitialSolutions});$   
   $list_{InitialSolutions} \leftarrow [ ];$   
**end**  
**for**  $i = 1$  **to**  $max = 100$  **do**  
   $map_{bests} \leftarrow Selection(map_{InitialPopulation});$   
   $map_{feasibles} \leftarrow Workable(map_{bests}, HL, CL);$   
  **for**  $key : map_{feasibles}.keySet()$  **do**  
     $list_{feasibleSolutions}.add(map_{feasibles}.get(key));$   
  **end**  
   $map_{crossover} \leftarrow Crossover(map_{bests});$   
   $map_{mutation} \leftarrow Mutation(map_{crossover}, HL, CL);$   
   $map_{NG} \leftarrow NewGeneration(map_{bests}, map_{mutation});$   
   $map_{InitialPopulation} \leftarrow map_{NG};$   
**end**  
 $list_{res} \leftarrow optimize(list_{feasibleSolutions}, HL, CL);$   
 Deploy list of feasible solutions ( $list_{res}$ ) that represent the best container placements in the host;

---

Algorithm 1 shows the principle of the genetic algorithm and the main methods used to achieve the objective of the problem. Input data is set before starting initialization. This entry will help the different operations related to the algorithm to determine the possible feasible solutions that will be the result, which represent several best container placements.

The algorithm starts by checking the selected input data, which are the hosts and the containers to be placed. This operation checks whether the total sum of Rams of all outbound hosts is greater than that of containers, to ensure that there is sufficient space for the placement of the container request. Then, if the verification is done and the input data is accepted, then the construction of the initial population begins by creating four chromosomes. Each individual built (as a list) will be added to the Map which represents the initial population until the end of the operation by obtaining an initial generation of four individuals.

The selection phase will be applied to the initial population obtained to determine the best individuals who will be the parent chromosomes. The best individuals (parents) selected will be the entry of the method responsible on the crossing to produce new individuals (the sons), the result of this method will be after the entry of mutation.

In the end, a new generation will be created, after the end of the mutation phase, which will be the population of the second iteration. The final list of feasible solutions will go through the optimization

phase to further verify the placement of the containers and improve it if possible.

### 3.1.2 First-Fit Decreasing

First-Fit Decreasing is one of the best-known algorithms for the classic problem of Bin Packing [31]. The FFD's strategy for placement is defined by three points. At first, we set the elements in descending order of size. Secondly, we put each item we get there in the oldest bin (opened the earliest) into which it fits (whenever an item fits the capacity of bin 1, put it there, otherwise, it fits into bag 2, if it fits). Thirdly, the opening of a new bag or bin is only done if the item does not fit into a bag that already contains something [30, 32]. For our problem, the items to put in the bin will be the containers, while the bags or bins will be the hosts.

The following algorithm shows the process of placing containers in servers with the FFD.

---

**Algorithm 2:** FFD-based container placement
 

---

**Data:** List of container requests and list of hosts  
**Result:** Deploy containers on hosts  
 Set the container list in decreasing order of RAM;  
 $HS \leftarrow$  The size of the host list;  
 $CS \leftarrow$  The size of the container list;  
**for**  $i = 0$  **to**  $CS - 1$  **do**  
  **for**  $j = 0$  **to**  $HS - 1$  **do**  
    The *Resource* is a vector that contains the values of four resources (RAM, CPU, Storage, Bandwidth);  
    **if**  $Resource_{host_j} \geq Resource_{container_i}$  **and**  
       $!PrevHostList.contains(i)$  **then**  
         $PlacementList.add(j);$   
         $Resource_{host_j} \leftarrow$   
           $Resource_{host_j} - Resource_{container_i};$   
         $PrevHostList.add(i);$   
      **end**  
    **end**  
  **end**  
 Deploy the *PlacementList* that contains the placement of the container list in the hosts;

---

## 3.2 Forecasting Energy Consumption

To predict the energy consumption in the data centers, we will calculate the energy consumed by the system of active hosts and storage. This energy will help us predict the energy consumed by other equipment such as the cooling system. The following figure shows the distribution of energy consumed by a data center based on studies and previous work.

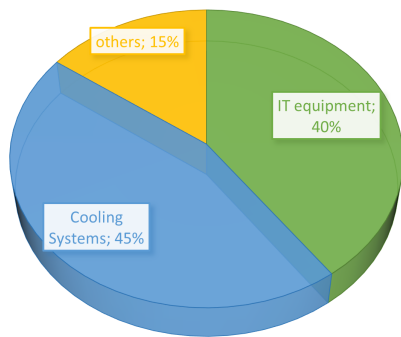


Figure 5: The distribution of energy consumption of a data center

To predict the energy consumed by active hosts, we rely on the placement of containers in these servers. This workload placement will help us determine the percentage used for hardware resources (RAM, CPU, Storage, Bandwidth) of each active host. The calculated percentages will be sorted between 0 (0%) and 1 (100%) to simplify their use in the CloudSim power model. Based on these percentages, we will define for each host the energy consumed by each resource in watts. The following equation represents the energy consumed in watts by a resource according to its percentage of use for a host.

$$Power_{resource_i} = getPower(Use_{resource_i}) \quad (3)$$

Predicting the energy consumed by each resource will help us to define the energy consumption for a single host and for all active servers that are defined by the following two equations.

$$Power_{host} = \sum_{i=1}^4 getPower(Use_i), Use_i \in [0, 1] \quad (4)$$

The  $Use_1$  is the used value of the RAM,  $Use_2$  is the used value of CPU,  $Use_3$  is the used value of the BW, and  $Use_4$  is the used value of the Storage.

$$Power_{AllHosts} = \sum_{i=1}^N Power_{host} \quad (5)$$

We represent below our algorithm for predicting the energy consumed by a data center based on container placement.

---

### Algorithm 3: Forecasting energy consumption

---

**Data:** List of container requests and list of hosts

**Result:** Predict the energy consumed by the data center  
Apply the Genetic Algorithm or FFD to predict container placement;

Group the percentages of use of each resource for each server in a map named *UseMap*;

**for** key : *UseMap.keySet()* **do**

**for** i = 0 to *UseMap.get(key).size()* **do**

        Apply the Equation (3) to calculate the energy consumed by each resource (RAM, CPU, Storage, Bandwidth); Apply Equation (4) to calculate the total energy consumed by a single host;

**end**

        Apply Equation (5) to predict the energy consumed by all active servers;

**end**

*DataCenterPower(Power\_{AllHosts});*

---

So, to predict the total energy consumed by a data center, we will rely on the energy predicted for the active host system. Algorithm (3) predicts the energy consumed by the active hosts in the data center based on the placement of workloads. Based on Figure 5 and previous work, we noted that the energy consumption of a data center is divided into three classes. The cooling system takes a large proportion with a value between 45% and 50% (47% on average). For the active host system and storage equipment, their consumption is between 36% and 40% (38% on average). Other systems such as lighting and communication equipment consume 15% of the total energy.

## 4 Experiments and Results

To evaluate our approach, we used CloudSim 4.0 to apply our algorithms to homogeneous and heterogeneous cloud systems. Our experiments will present the prediction of container placement, then the energy consumption for each system. In this section, we will perform three different experiments. In the first two applications, we will have two scenarios.

### 4.1 Application 1 : Homogeneous System

The homogeneous system we used in this experiment consists of 50 identical servers in terms of material resources and 3000 containers of different classes.

Table 1: Details of the Application 1

System	Number	RAM (MB)	Number of CPUs	Bandwidth	Storage (MB)
Hosts	50	65536	512	2000000	2000000
Containers (class 1)	1000	512	1	2500	1024
Containers (class 2)	1000	256	1	2500	1024
Containers (class 3)	1000	128	1	2500	1024

4.1.1 SCENARIO 1 : When all hosts are active and the percentage of use of each resource = 50%

In this scenario, we decided to predict the energy consumed by this data center in the case where the use of each material resource equals 50%. The following table shows the results obtained in detail.

Table 2: Energy consumption for the scenario 1 of Application 1

RAM use (watts)	CPU use (watts)	Bandwidth use (watts)	Storage use (watts)
5800	5800	5800	5800
Number of active hosts : 50			
Energy consumption of active hosts (38%)	Energy consumption of cooling system (47%)	Energy consumption of others (15%)	Energy consumption of the data center
23200.0 watts	28694.736 watts	9157.895 watts	61052.63 watts

4.1.2 SCENARIO 2 : Genetic Algorithm application

This scenario represents the application of our genetic algorithm for the prediction of container placement and the energy consumption of the different systems of the data center. The results obtained are as follows.

Table 3: Energy consumption for the scenario 2 of Application 1

RAM use (watts)	CPU use (watts)	Bandwidth use (watts)	Storage use (watts)
2288.932	1920.636	1828.676	1738.72
Number of active hosts : 18			
Energy consumption of active hosts (38%)	Energy consumption of cooling system (47%)	Energy consumption of others (15%)	Energy consumption of the data center
7776.964 watts	9618.877 watts	3069.8542 watts	20465.693 watts

4.1.3 SCENARIO 3 : FFD application

In this scenario, we applied the First-Fit Decreasing algorithm to place the container list and predict the energy consumed by this data center after completing the workload placement. The following table shows the results obtained.

Table 4: Energy consumption for the scenario 3 of Application 1

RAM use (watts)	CPU use (watts)	Bandwidth use (watts)	Storage use (watts)
2288.932	1826.91	1830.826	1739.208
Number of active hosts : 18			
Energy consumption of active hosts (38%)	Energy consumption of cooling system (47%)	Energy consumption of others (15%)	Energy consumption of the data center
7782.606 watts	9625.8545 watts	3072.0813 watts	20480.543 watts

The objective of the first scenario is to see the rate of energy that will be consumed if a data center hosts a given workload but without any strategy to define its conception so as not to fall into the problem of waste of resources which implies an expansion of energy consumption. In this scenario, each host consumed 464 watts because they are identical. The fact that the 50 hosts are all active, involved an expansion of the total energy consumed that exceeded 61052 watts, which is ordinary because all servers are

active. This scenario shows the importance of defining a strategy for the placement of data center workloads and seeing the impact when a strategy is applied to place containers in an minimal number of servers. The application of the genetic algorithm gave two solutions of 18 servers for each that will be active to host the 3000 containers. We noticed in each solution that the power consumption of the hosts is close to each other (between 400 and 476 watts). As well as workloads are well distributed among active servers (a good load balancing). For total energy consumption, the first solution predicted a value of 20465.693 watts with 40586.937 saved energy compared to the first scenario. The second solution estimates the energy consumption with 20466.234 watts and 40586.396 watts of energy saved when compared with the solution of scenario 1.

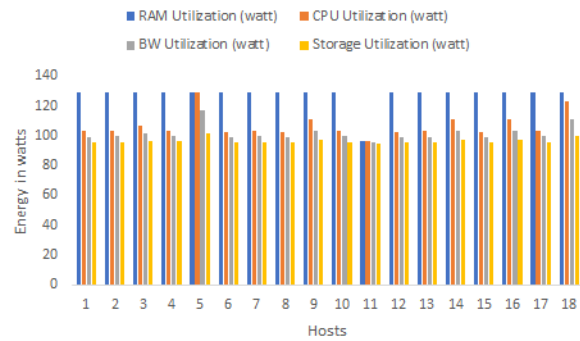


Figure 6: The energy consumed in watts by each resource for the GA first solution - Application 1

Like the genetic algorithm, the FFD proposed a single arrangement of 18 hosts among the starting 50 to place the list of containers. So, 32 will be retained, contrary to the first scenario. Of the 18 active servers, 9 hosts each consumed more than 423 watts, 8 others consumed between 427 and 476 watts for each, and only one server that is the last one consumed less than 400 watts (383.59146 watts). The placement proposed by the FFD also gave good results for the total energy consumption (20480.543 watts) which is lower than in the first scenario with a value of 40572.087 watts of energy saved. But the solutions of the genetic algorithm are better in terms of the energy consumed but identical to that of the FFD in the number of active hosts.

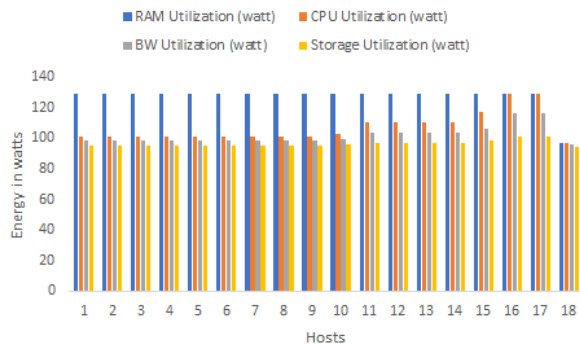


Figure 7: The energy consumed in watts by each resource for the FFD solution - Application 1

## 4.2 Application 2 : Heterogeneous System

In this application, we used a heterogeneous system that consists of three different types of hosts with a total number of 50, and 3000 containers.

Table 5: Details of the Application 2

System	Number	RAM (MB)	Number of CPUs	Bandwidth	Storage (MB)
Hosts (class 1)	17	32768	256	2000000	2000000
Hosts (class 2)	17	65536	512	2000000	2000000
Hosts (class 3)	16	131072	640	2000000	2000000
Containers (class 1)	1000	512	1	2500	1024
Containers (class 2)	1000	256	1	2500	1024
Containers (class 3)	1000	128	1	2500	1024

### 4.2.1 SCENARIO 1 : When all hosts are active and the percentage of use of each resource = 50%

In the first scenario of application 2, we predicted the energy consumed for a heterogeneous system when all servers are active.

Table 6: Energy consumption for the scenario 1 of Application 2

RAM use (watts)	CPU use (watts)	Bandwidth use (watts)	Storage use (watts)
6202	6202	6202	6202
Number of active hosts : 50			
Energy consumption of active hosts (38%)	Energy consumption of cooling system (47%)	Energy consumption of others (15%)	Energy consumption of the data center
24808.0 watts	30683.578 watts	9792.632 watts	65284.21 watts

### 4.2.2 SCENARIO 2 : Genetic Algorithm application

Scenario 2 represents the application of the genetic algorithm on our heterogeneous system. The results are as follows.

Table 7: Energy consumption for the scenario 2 of Application 1

RAM use (watts)	CPU use (watts)	Bandwidth use (watts)	Storage use (watts)
1944.637	1594.332	1450.27	1193.698
Number of active hosts : 11			
Energy consumption of active hosts (38%)	Energy consumption of cooling system (47%)	Energy consumption of others (15%)	Energy consumption of the data center
6182.937 watts	7647.3164 watts	2440.633 watts	16270.887 watts

### 4.2.3 SCENARIO 3 : FFD application

For this scenario, we applied the FFD algorithm. Below is the prediction of the energy consumed by our system after this application.

Table 8: Energy consumption for the scenario 3 of Application 2

RAM use (watts)	CPU use (watts)	Bandwidth use (watts)	Storage use (watts)
3049.891	2592.923	2448.035	2358.184
Number of active hosts : 26			
Energy consumption of active hosts (38%)	Energy consumption of cooling system (47%)	Energy consumption of others (15%)	Energy consumption of the data center
10449.033 watts	12923.805 watts	4124.6187 watts	27497.457 watts

In the second application, we used a heterogeneous system to see if the placement of the containers and the rate of energy consumption will be influenced by the nature of the system. In the first scenario, the 50 hosts consumed a different energy rate because they are heterogeneous. The first 17 servers consumed 408 watts each, and the last 16 consumed the high value (624 watts each). The other servers consumed 464 watts each. For the total energy consumption, the system consumed 65284.21 watts with superiority of 4231.58 compared to scenario 1 of the first application.

The application of the genetic algorithm for this heterogeneous system gave two different solutions in contrast to scenario 2 of Experiment 1 which proposed two identical solutions at the level of the proposed number of active hosts. The heterogeneous nature of this system has increased the chance of having diversified solutions in terms of the hosts used and their number.

The first solution obtained by the genetic algorithm has proposed 13 servers that will be active to host the container list, which is the best number compared to the first scenario. The energy consumption of the 13 active servers is between 350 and 800 watts because of the hosts' diversity in the hardware resources. This diversity influenced the rate of energy consumed by the data center which did not exceed 17623.83 watts with 47660.38 watts of energy saved compared to scenario 1.

The second solution was the best with 11 active hosts and less than 2 servers compared to the first solution to host workloads. The values consumed of energies differ from one host to another because of the diversity in material resources. This number of active servers influenced the energy consumption rate, which was minimized at the level of all the data center systems and 1352.943 watts for the total consumption.

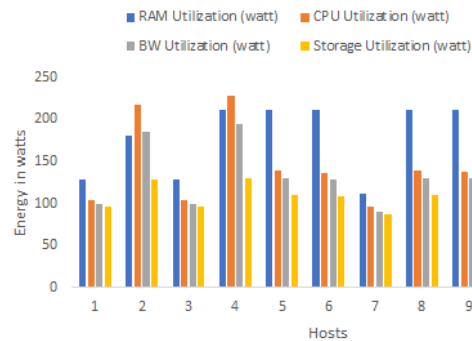


Figure 8: The energy consumed in watts by each resource for the GA second solution - Application 2

The unique solution of the FFD is different from those of the



genetic algorithm. The solution offers 26 servers for container placement which is not optimal when compared with the solutions of scenario 2. This increase in the number of active hosts for the FFD is because of its way of assigning containers to servers which is classic and avoids waste of space but is not always effective if there is a list of containers in a defined order (increasing or decreasing). For this reason, most of the active hosts (17) belong to class 1 and the others belong to the second class.

So workloads were not well distributed between the active hosts about the solution of the genetic algorithm, which is a weak point in the FFD algorithm. The high number of active servers consumed between 370 and 480 watts for each, implied an increase in the rate of energy consumed by the data center with a value of 27497.457 watts.

We note that the genetic algorithm guarantees best container placement and minimal energy consumption due to its functioning which is adaptive to different types of systems. But in general, both algorithms offer a good solution for virtualization in a data center, which is important for the whole system to operate without any waste of resources that can increase energy consumption

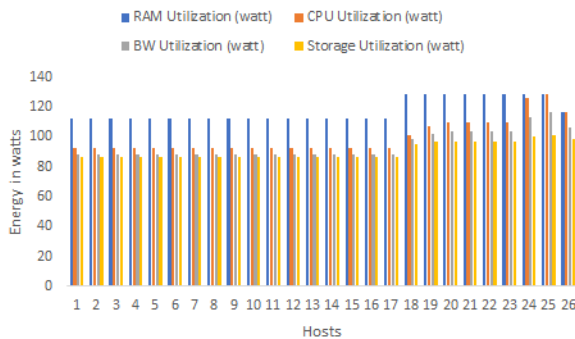


Figure 9: The energy consumed in watts by each resource for the FFD solution - Application 2

### 4.3 Application 3 : Global Comparison

After testing our algorithms, we decided to compare them with other previous heuristic or meta-heuristic methods to properly evaluate our algorithms. For this, we chose the approach in [19] which uses the Simulated Annealing (SA) algorithm to predict the placement of virtual machines based on the estimate of energy consumption to choose the best placement. In our case, we will evaluate this algorithm using container instances to compare it with our genetic algorithm (GA). More of this we will compare the approach proposed in [33] which applies the Ant Colony Optimization (ACO) for container placement with our genetic algorithm (GA).

The choice of the Simulated Annealing and ACO to examine the effectiveness of our genetic algorithm was not random, but due to several reasons. First of all these three algorithms belong to the class of metaheuristics which is inspired by nature to solve optimization problems. More of that, they can adapt to problems with a high complexity and which require a huge calculation. Another point that encouraged us more to compare our Genetic Algorithm with Simulated Annealing and the ACO is that the works [19] and [33] have

the same vision regarding the placement of virtual instances by optimizing the use of material resources and uses the same procedure, except that the work [19] are based on virtual machines instead of containers. For the FFD, we will evaluate it with First-Fit (FF) and Random-Fit (RF) which are algorithms of the Bin Packing problem such as the FFD. Table 9 shows the resources of the different cloud systems used on which we will perform our comparison.

Table 10 represents the results obtained for the energy consumption of several data centers for each algorithm.

Table 10 represents three main results for six algorithms. For each system, we applied six algorithms to predict the energy consumed by each data center based on container placement. As well as, we calculated the energy saved or conserved for each system by comparing the predicted energy consumption with the energy consumed when all the hosts of a system are active and the percentage of use of each resource = 50%. In addition, we calculated the execution time of each algorithm.

For the first system which is homogeneous, we notice that the energy consumption obtained by the genetic algorithm does not exceed 13460 watts with 47594.73 watts of saved energy. For the ACO algorithm, the energy consumption exceeded the value predicted by the genetic algorithm twice with 32931.009 watts of stored energy. Similarly, for the SA algorithm, the energy consumed exceeded that of GA but with a big difference which influenced the rate of energy saved which was less than 10380 watts. On the other hand, the execution time of the genetic algorithm was very high (more than 31 seconds) compared to that of ACO which gave its result in 0.297 seconds and the SA in 5.62 seconds. The reason why the execution time of the genetic algorithm is very high is because of its execution process that operates under several iterations (in our case 100 iterations).

For heuristics, the three Bin Packing problem-solving algorithms (FFD, FF, and RF) predicted energy consumption values close to each other. The values proposed by the FFD and FF were identical to the superiority of the RF algorithm which gave a value of 13447.468 watts which is optimal compared to those of the FFD and FF. For the energy saved, the three heuristics provided better values with the genetic algorithm, but they are fast in execution time because they produce a single solution without using several iterations.

In the second system, which is heterogeneous, the results obtained by the genetic algorithm for the energy consumed and saved are the best when compared with those of ACO and SA and also with the three heuristics. On the other hand, its execution time which reached 31 seconds is very high compared to the other algorithms. We also notice that the FFD, FF, and RF gave the same energy consumption value (31874.584 watts), but the RF surpasses them in the execution time (0.08 seconds). The same goes for the third system, which has a superiority of the genetic algorithm over the others in terms of the energy consumed which is optimal, but the execution time has increased greatly (301.769 seconds) because of the size of the number of instances in this system. The SA also took 103.123 seconds which is high compared to other methods.

For other systems, our genetic algorithm was the best in systems 5 and 6 in terms of optimal energy consumption, but it takes a long time to finish its execution and give the final results. In system 4, the FFD and FF heuristics proposed optimal solutions with an energy

Table 9: Details of the Application 3

System	Hosts					Containers				
	Number	RAM (GB)	Number of CPUs	Bandwidth	Storage (MB)	Number	RAM (MB)	Number of CPUs	Bandwidth	Storage (MB)
1	50	32	256	2000000	2000000	334	512	1	1024	2500
						334	256	1	1024	2500
						332	128	1	1024	2500
2	17	16	128	2000000	2000000	1000	512	1	1024	2500
	17	32	256	2000000	2000000					
	16	64	512	2000000	2000000					
3	25	16	128	2000000	2000000	667	128	1	1024	2500
	25	32	256	2000000	2000000	667	256	1	1024	2500
	25	64	512	2000000	2000000	666	512	1	1024	2500
4	100	64	512	2000000	2000000	1667	512	1	1024	2500
						1667	256	1	1024	2500
						1666	128	1	1024	2500
5	34	32	256	2000000	2000000	1667	128	1	1024	2500
	34	64	512	2000000	2000000	1667	256	1	1024	2500
	32	128	1024	2000000	2000000	1666	512	1	1024	2500
6	30	256	1024	4000000	4000000	3500	256	1	1024	2500
	30	128	1024	4000000	4000000					
	30	64	512	4000000	4000000	3500	512	1	1024	2500
	30	32	256	4000000	4000000					

Table 10: Comparison of different algorithms for predicting the energy consumption of different cloud systems

System	GA			ACO			SA		
	Data Center Energy (watt)	Energy Saved (watt)	Execution time (second)	Data Center Energy (watt)	Energy Saved (watt)	Execution time (second)	Data Center Energy (watt)	Energy Saved (watt)	Execution time (second)
1	13457.9	47594.73	31.954	28121.621	32931.009	0.297	50678.336	10374.294	5.62
2	18792.02	42260.61	31.004	39339.375	21713.255	0.179	51927.41	9125.22	4.034
3	20413.207	71165.743	301.769	53369.266	38209.684	0.264	77419.08	14159.87	103.123
4	43210.434	121000.086	5304.5	67590.805	96619.715	1.496	100156.78	64053.74	3870.694
5	28867.098	101701.322	3886.164	77247.85	53320.57	1.17	98288.49	32279.93	1165.504
6	19218.874	148149.546	11634.408	31827.826	135540.594	1.126	116079.95	51288.47	1317.876

System	FFD			FF			RF		
	Data Center Energy (watt)	Energy Saved (watt)	Execution time (second)	Data Center Energy (watt)	Energy Saved (watt)	Execution time (second)	Data Center Energy (watt)	Energy Saved (watt)	Execution time (second)
1	13462.325	47590.305	0.194	13462.325	47590.305	0.122	13447.468	47605.162	0.118
2	31874.584	29178.046	0.116	31874.584	29178.046	0.062	31874.584	29178.046	0.08
3	40164.184	51414.766	0.351	40546.266	51032.684	0.2	40333.258	51245.692	0.287
4	43202.652	121007.868	7.788	43202.652	121007.868	8.441	43329.03	120881.49	2.856
5	48174.152	82394.268	2.93	48301.848	82266.572	2.594	48233.105	82335.315	2.186
6	21229.072	167368.42	12.186	21242.38	146126.04	8.91	21098.488	146269.932	10.894

value of 43202.652 watts which is less than 7782 watts compared to the value proposed by the genetic algorithm. Generally, our GA has proposed best values for energy consumption for most systems ahead of other metaheuristics (ACO and SA) and even heuristics. On the other hand, execution time remains its main weakness because of its way of solving the problem. FFD was best with RF in the heuristics used. In terms of execution time, the ACO was the best with an average time of 0.75 seconds. More of this the SA algorithm proposed large energy values and ranked second before

the genetic algorithm at the level of execution time.

## 5 Conclusion

In this paper, we have presented an approach based on heuristics and metaheuristics for predicting the energy consumed by different data centers based on the placement of workloads. Our approach has proposed a genetic algorithm and a First-Fit Decreasing using

cloud containers to predict the best and optimal placement for these instances in several servers without wasting hardware resources. The results obtained showed that the genetic algorithm guarantees a good placement for containers and minimizes the energy consumption in the data centers, after comparing it with other metaheuristics such as Ant Colony Optimization and Simulated Annealing.

**Conflict of Interest** The authors declare no conflict of interest.

## References

- [1] A. Bouaouda, K. Afdel, R. Abounacer, "Forecasting the Energy Consumption of Cloud Data Centers Based on Container Placement with Ant Colony Optimization and Bin Packing," in 2022 5th Conference on Cloud and Internet of Things (CIoT), 150–157, 2022, doi:10.1109/CIoT53061.2022.9766522.
- [2] A. Greenberg, J. Hamilton, D. A. Maltz, P. Patel, "The Cost of a Cloud: Research Problems in Data Center Networks," SIGCOMM Comput. Commun. Rev., **39**(1), 68–73, 2009, doi:10.1145/1496091.1496103.
- [3] G. Wu, M. Tang, Y.-C. Tian, W. Li, "Energy-Efficient Virtual Machine Placement in Data Centers by Genetic Algorithm," in T. Huang, Z. Zeng, C. Li, C. S. Leung, editors, *Neural Information Processing*, 315–323, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
- [4] A. Mosa, N. Paton, "Optimizing virtual machine placement for energy and SLA in clouds using utility functions," *Journal of Cloud Computing*, **5**, 2016, doi:10.1186/s13677-016-0067-7.
- [5] "Cloud Computing Energy Efficiency, Strategic and Tactical Assessment of Energy Savings and Carbon Emissions Reduction Opportunities for Data Centers Utilizing SaaS, IaaS, and PaaS," Technical report, Pike Research, 2010.
- [6] M. Dayarathna, Y. Wen, R. Fan, "Data Center Energy Consumption Modeling: A Survey," *IEEE Communications Surveys Tutorials*, **18**(1), 732–794, 2016, doi:10.1109/COMST.2015.2481183.
- [7] "Energy efficiency policy options for australian and new zealand data centres," Technical report, The Equipment Energy Efficiency (E3) Program, 2014.
- [8] R. Brown, E. Masanet, B. Nordman, B. Tschudi, A. Shehabi, J. Stanley, J. Koomey, D. Sartor, P. Chan, J. Loper, S. Capana, B. Hedman, "Report to Congress on Server and Data Center Energy Efficiency: Public Law 109-431," 2007, doi:10.2172/929723.
- [9] D. Boru, D. Kliazovich, F. Granelli, P. Bouvry, A. Y. Zomaya, "Energy-efficient data replication in cloud computing datacenters," in 2013 IEEE Globecom Workshops (GC Wkshps), 446–451, 2013, doi:10.1109/GLOCOMW.2013.6825028.
- [10] W. Van Heddeghem, S. Lambert, B. Lannoo, D. Colle, M. Pickavet, P. Demeester, "Trends in worldwide ICT electricity consumption from 2007 to 2012," *Computer Communications*, **50**, 64–76, 2014, doi:https://doi.org/10.1016/j.comcom.2014.02.008, green Networking.
- [11] "Facts & stats: Data architecture and more data," Technical report, Info-Tech, 2010.
- [12] Greenpeace, "Make It Green: Cloud Computing and its Contribution to Climate Change," Technical report, Greenpeace International, 2010.
- [13] A. Sharma, N. Nitin, "A Multi-Objective Genetic Algorithm for Virtual Machine Placement in Cloud Computing," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, **8**, 2278–3075, 2019.
- [14] G. Dasgupta, A. Sharma, A. Verma, A. Neogi, R. Kothari, "Workload Management for Power Efficiency in Virtualized Data Centers," *Commun. ACM*, **54**(7), 131–141, 2011, doi:10.1145/1965724.1965752.
- [15] G. Warkozek, E. Drayer, V. Debusschere, S. Bacha, "A new approach to model energy consumption of servers in data centers," in 2012 IEEE International Conference on Industrial Technology, 211–216, 2012, doi:10.1109/ICIT.2012.6209940.
- [16] L. Liu, H. Wang, X. Liu, X. Jin, W. He, Q. Wang, Y. Chen, "GreenCloud: a new architecture for green data center," 2009, doi:10.1145/1555312.1555319.
- [17] X. Fan, W.-D. Weber, L. A. Barroso, "Power Provisioning for a Warehouse-Sized Computer," *SIGARCH Comput. Archit. News*, **35**(2), 13–23, 2007, doi:10.1145/1273440.1250665.
- [18] D. Meisner, T. F. Wenisch, "Peak power modeling for data center servers with switched-mode power supplies," in 2010 ACM/IEEE International Symposium on Low-Power Electronics and Design (ISLPED), 319–324, 2010, doi:10.1145/1840845.1840911.
- [19] K. Dubey, S. C. Sharma, A. A. Nasr, "A Simulated Annealing based Energy-Efficient VM Placement Policy in Cloud Computing," in 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 1–5, 2020, doi:10.1109/ic-ETITE47903.2020.119.
- [20] Y. Wu, M. Tang, W. Fraser, "A simulated annealing algorithm for energy efficient virtual machine placement," in 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 1245–1250, 2012, doi:10.1109/ICSMC.2012.6377903.
- [21] J. Xu, J. A. B. Fortes, "Multi-Objective Virtual Machine Placement in Virtualized Data Center Environments," in 2010 IEEE/ACM Int'l Conference on Green Computing and Communications and Int'l Conference on Cyber, Physical and Social Computing, 179–188, 2010, doi:10.1109/GreenCom-CPSCom.2010.137.
- [22] Y. Gao, H. Guan, Z. Qi, Y. Hou, L. Liu, "A multi-objective ant colony system algorithm for virtual machine placement in cloud computing," *Journal of Computer and System Sciences*, **79**(8), 1230–1242, 2013, doi:https://doi.org/10.1016/j.jcss.2013.02.004.
- [23] S. Pang, W. Zhang, M. Tongmao, Q. Gao, "Ant Colony Optimization Algorithm to Dynamic Energy Management in Cloud Data Center," *Mathematical Problems in Engineering*, **2017**, 1–10, 2017.
- [24] A. Kansal, F. Zhao, J. Liu, N. Kothari, A. A. Bhattacharya, "Virtual Machine Power Metering and Provisioning," in *Proceedings of the 1st ACM Symposium on Cloud Computing, SoCC '10*, 39–50, Association for Computing Machinery, New York, NY, USA, 2010, doi:10.1145/1807128.1807136.
- [25] P. Dziuranski, S. Zhao, M. Przewozniczek, M. Komarnicki, L. S. Indrusiak, "Scalable distributed evolutionary algorithm orchestration using Docker containers," *Journal of Computational Science*, **40**, 101069, 2020, doi:https://doi.org/10.1016/j.jocs.2019.101069.
- [26] C. Zaher, "For CTO's: the no-nonsense way to accelerate your business with containers," Technical report, Canonical Limited 2017. Ubuntu, Kubuntu, 2017.
- [27] D. Bernstein, "Containers and Cloud: From LXC to Docker to Kubernetes," *IEEE Cloud Computing*, **1**(3), 81–84, 2014, doi:10.1109/MCC.2014.51.
- [28] O. Kramer, *Genetic Algorithms*, 11–19, Springer International Publishing, Cham, 2017, doi:10.1007/978-3-319-52156-5\_2.
- [29] I. Boussaïd, J. Lepagnot, P. Siarry, "A survey on optimization metaheuristics," *Information Sciences*, **237**, 82–117, 2013, doi:https://doi.org/10.1016/j.ins.2013.02.041, prediction, Control and Diagnosis using Advanced Neural Computations.
- [30] N. Janani, R. Jegan, P. Prakash, "Optimization of Virtual Machine Placement in Cloud Environment Using Genetic Algorithm," *Research Journal of Applied Sciences, Engineering and Technology*, **10**, 274–287, 2015, doi:10.19026/rjaset.10.2488.
- [31] A. Wolke, B. Tsend-Ayush, C. Pfeiffer, M. Bichler, "More than bin packing: Dynamic resource allocation strategies in cloud data centers," *Information Systems*, **52**, 83–95, 2015, doi:https://doi.org/10.1016/j.is.2015.03.003, special Issue on Selected Papers from SISAP 2013.
- [32] R. Panigrahy, K. Talwar, L. Uyeda, U. Wieder, "Heuristics for Vector Bin Packing," 2011.
- [33] O. SMIMITE, K. AFDEL, "Hybrid Solution for Container Placement and Load Balancing based on ACO and Bin Packing," *International Journal of Advanced Computer Science and Applications*, **11**(11), 2020, doi:10.14569/IJACSA.2020.0111174.

## Characterization and Investigating the Effect of Gate-Insulator Thickness on Co-Axial Cylindrical Carbon Nanotube Field Effect Transistor

Suchismita Sen<sup>1</sup>, Argha Sarkar<sup>2\*</sup>, Pinaki Chakraborty<sup>1</sup>

<sup>1</sup>Department of Physics, Raiganj University, Raiganj, 733134, India

<sup>2</sup>School of Computer Science and Engineering, REVA University, Bangalore, 560064, India

### ARTICLE INFO

Article history:

Received: 29 August, 2022

Accepted: 12 December, 2022

Online: 24 January, 2023

Keywords:

Carbon nanotube

FETToy

Carbon nano tube diameter

Characterization curve

### ABSTRACT

Carbon nanotube field effect transistor (CNTFET) has a huge advantage over the Si-MOSFET. In MOSFET switching occurs by altering channel resistivity whereas in CNTFET switching occurs by modulation contact resistance. CNTFET generates three to four times of drive current than MOSFET. Transconductance of CNTFET is four times higher than the MOSFET. The average carrier velocity is also very high almost double in CNTFET than that is in MOSFET. Its power consumption is low. Electron mobility is high. Threshold voltage is also low. It has better control over channel formation. There is no direct tunneling and gate leakage current is also reduced. Herein, the main objective is to investigate the effect of gate-insulator thickness on CNTFET, and to optimize the thickness so that current carrying capacity may reach higher. A detailed simulations have been made and IV characterization is done to investigate the effect of Gate-Insulator Thickness on Co-Axial Cylindrical Carbon Nanotube Field Effect Transistor. Report shows with the increasing gate-insulator thickness current is decreased significantly. Where as the variation of nano diameter shows that the increasing rate of current is increased when the carbon tube diameter is increased.

### 1. Introduction

In this age of nanotechnology, the demand of integrated circuits with smaller dimension has increased. On the other side this fast-moving world requires technology with high speed performance and that can consume lower power. To fulfill such requirements, the use of carbon nanotube field effect transistor (CNTFET) over Si-MOSFET has increased widely. Its ability to carry high current makes it more popular [1]. Carbon nanotubes consist of carbon atoms having diameter in nanometer range [2]. The considerable tiny sized carbon and its electronic configuration ensure unique carbon element with versatile structures and alluring properties [2]. Having the title of the strongest material ever measured, graphene is a two-dimensional (one-atom-thickness) allotrope of carbon with a planar honeycomb lattice [3]. It is regarded as the basic building block of carbon nanotubes. The versatile properties of carbon nanotubes (CNT) basically sourced from graphene [2]. Folding one or multiple graphene sheets with a specific chiral angle creates unique CNT.

Based on the number of folded layers' carbon nanotubes can be classified in two types.

- Single – walled carbon nanotube (SWNTs), having diameter 1nm [4]
- Multi – walled carbon nanotube (MWNTs), having diameter 100 nm

In multilayer formation many layers are interlinked. On the other hand, another classification of CNTFETs can be mentioned based on its geometry.

In a back-gate CNTFET generally SWCNT is used. It was first proposed by Tans et.al. [5]. The I(on)/ I(off) ratio of this type of CNTFET is almost 105 [6]. The parasitic contact resistance of such CNTFET is very high (>1Mohm) [7]. On the other hand, the drain current as well as the value of transconductance is very low. Drain current is of the nano range [6]. Such limitations of back-gate CNTFET drive the researchers to develop a next generation CNTFET.

\*Corresponding Author: Argha Sarkar, Email: argha15@gmail.com

Wind et al. have come up with the first top gate CNTFET [8]. In this model the gate is formed over the carbon nanotubes. Though the fabrication process of Top gate CNTFET is little complicated but it is preferred over back-gate CNTFET due to its high drain current of the order of micro and for the greater value of transconductance.

Unlike the other two CNTFETs in Wrap around gate CNTFET the whole nanotube is covered by gate. It is also known as Gate-all-around CNTFET. To expose the ends of the nanotube the wrapping is partially etched and then the source, gate and drain contacts are deposited on the nanotubes. As the entire carbon nanotube is covered, it reduces the leakage current and increases the electrical performance.

In suspended CNTFET method, gate is suspended over a trench to reduce the contact with substrate and gate oxide and it improves the device performance. But the main drawback of such type of CNTFET is here air or vacuum is considered as the dielectric medium. Only short CNTs are used as long tubes may short the device by touching the metal contact.

Depending on the type of electrodes used, the CNTFET classification has been made into three categories. (a) Schottky-barrier (SB) CNTFET (b) Partially gated (PG) CNTFET and (c) doped-S/D CNTFET [9-16]. And the differences are clearly shown in Figure 1.

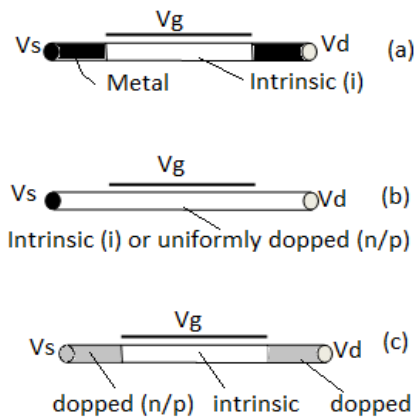


Figure 1: Different types of CNTFET: (a) Schottky-barrier (SB) CNTFET, (b) partially gated (PG) CNTFET (c) doped-S/D CNTFET [9].

Here we have focused on the co-axial cylindrical CNTFET. In such CNTFETs an oxide layer is portrayed around a cylindrical carbon nano tube. Further a metallic cylindrical layer is deposited in it. This metallic contact can behave as a gate here. At a fixed bias voltage it can induce more channel charge than the other CNTFETs. This is because of its geometry. The capacitive coupling between the gate electrode and the nanotube surface is the maximum for it. Technologies like complementary metal-oxide-semiconductor (CMOS) can be affected by the short channel effects. This improved coupling can prevent this short channel effects. Its geometry of end contact is also important as it can provide the concept of the dimension of Schottky barrier. This Schottky barrier is actually present at the channel near device ends and it can directly influence the current modulation. It also has a huge role in low voltage applications.

## 2. Simulation of Co-Axial Cylindrical Carbon Nanotube Field Effect Transistor

The In this paper simulations were done for co – axial cylindrical CNTFET using the well-known FETToy tool to see how the characteristic curves depends on different tube parameters like nanotube diameter and gate insulator thickness. Varying the Gate Oxide thickness and nano-tube diameter the drain current can be varied. On the other hand the scaling of the most popular Si-MOSFET almost approaches towards its limiting values. In search of new alternatives this simulation was done to overcome these limitations.

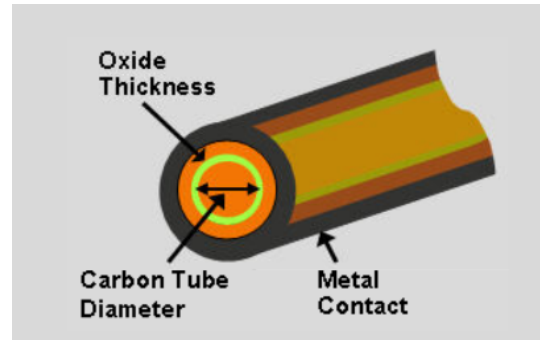


Figure 2: co-axial cylindrical CNTFET

Figure 2 represents a schematic diagram of co-axial cylindrical CNTFET. First for a constant nanotube diameter (1 nm), the simulation was done by varying the gate insulator thickness only. For this simulation the ambient temperature was taken as 300K. Threshold voltage of the used CNTFET was 0.32 V whereas the gate control parameter and drain control parameter were considered as 0.88 and 0.35 respectively. Series resistance was taken zero.

## 3. Prepare Your Paper before Styling

Before Some data taken during the simulation are given below.

Table 1: (Drain voltage=1 V, Nanotube diameter=1nm)

Gate Voltage (V)	Drain Current ( $\mu$ A)		
	Gate-Insulator thickness 1nm	Gate-Insulator thickness 1.5nm	Gate-Insulator thickness 2nm
0	0.00672	0.0000661	0.0000661
0.0833	0.114	0.00112	0.00112
0.1667	1.9	0.0187	0.0186
0.2500	24.4	0.24	0.232
0.3333	123	1.23	1.13
0.4167	296	3.01	2.69
0.5000	517	5.41	4.75
0.5833	778	8.37	7.26
0.6667	1070	11.9	10.2
0.7500	1400	15.9	13.7
0.8333	1750	20.4	17.5
0.9167	2140	25.1	21.6
1.0000	2550	30	26

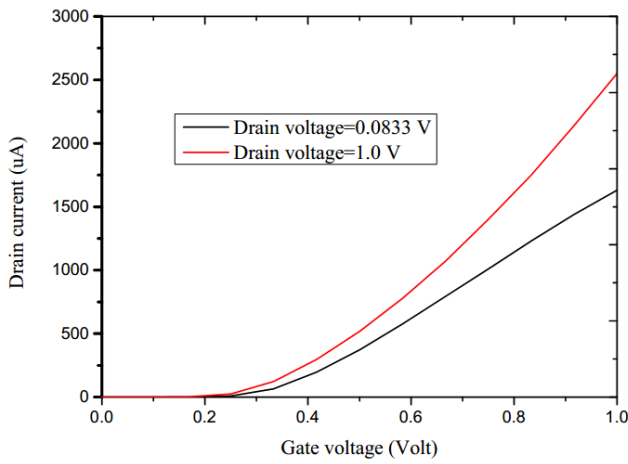


Figure 3: (a) Gate Voltage Vs Drain Current curve for a CNT having diameter 1nm and gate insulator thickness 1nm.

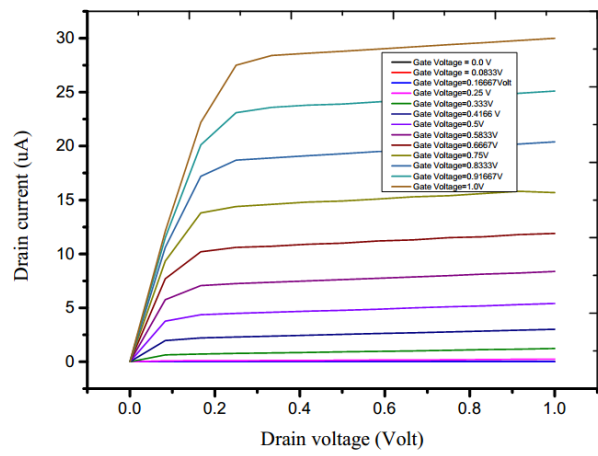


Figure 3: (d) Output Characteristic curve for a CNT having diameter 1nm and gate insulator thickness 1.5nm.

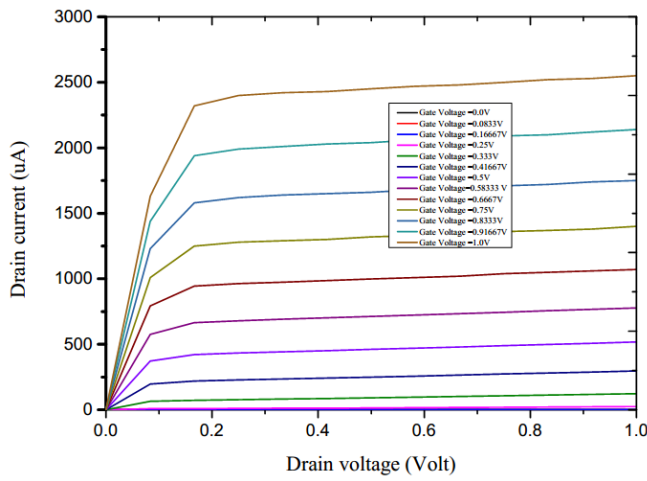


Figure 3: (b) Output Characteristic curve for a CNT having diameter 1nm and gate insulator thickness 1nm .

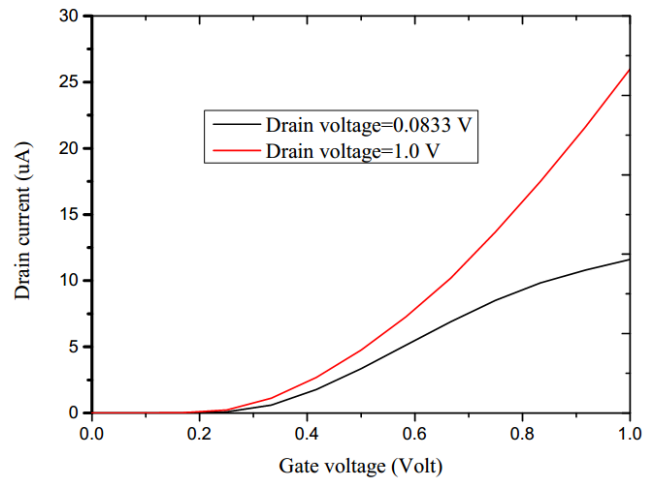


Figure 3: (e) Gate Voltage Vs Drain Current curve for a CNT having diameter 1nm and gate insulator thickness 2nm.

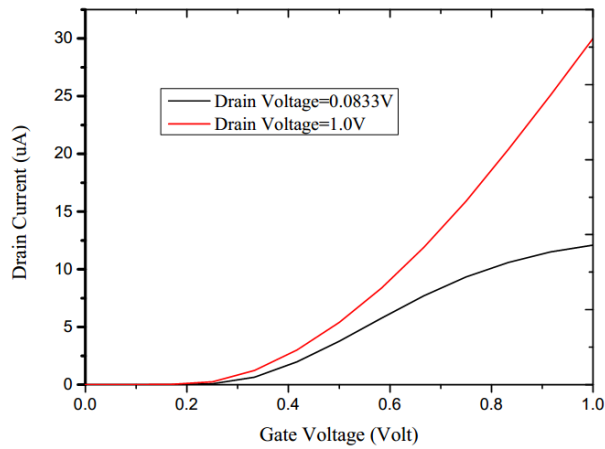


Figure 3: (c) Gate Voltage Vs Drain Current curve for a CNT having diameter 1nm and gate insulator thickness 1.5nm.

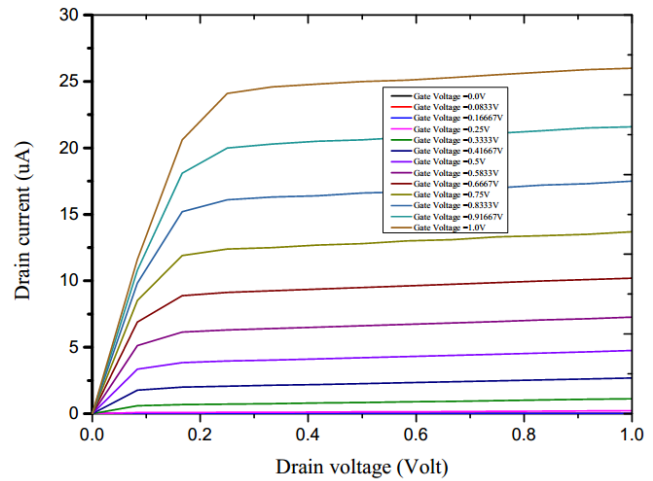


Figure 3: (f) Output Characteristic curve for a CNT having diameter 1nm and gate insulator thickness 2nm.

From the data of Table-1 it is clear that the drain current vs gate voltage characteristic curve for 1nm gate-insulator thickness is steeper than the other two. Whereas for the same given voltage CNTFET having gate-insulator thickness 2nm gives the lowest drain current. So the drain current is decreased with the increasing gate-insulator thickness. This is due to the fact that with the increase of the gate-insulator thickness the resistance across it is also increased and as a result the drain current is decreased [12]. When the gate dielectric becomes thicker, the electric field within the dielectric becomes smaller for the same gate voltages. Thus the accumulated free carrier near the interface also becomes less. As the carrier density decreases, the drain current decreases as well. Figure 3 shows the input and output characteristic curves for CNTs having different gate insulator thickness but same nanotube diameter i.e. 1nm. Another simulation (shown in figure 4) was done by taking nanotube diameter as 2 nm and varying the gate-insulator thickness from 1nm to 2nm. Other aspects were fixed. I- V characterization is made to investigate the Effect of gate-insulator thickness on co-axial cylindrical CNTFET.

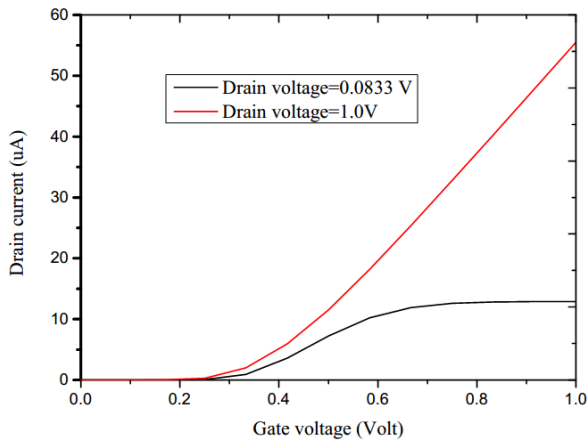


Figure 4: (a) Gate Voltage Vs Drain Current curve for a CNT having diameter 2nm and gate insulator thickness 1nm

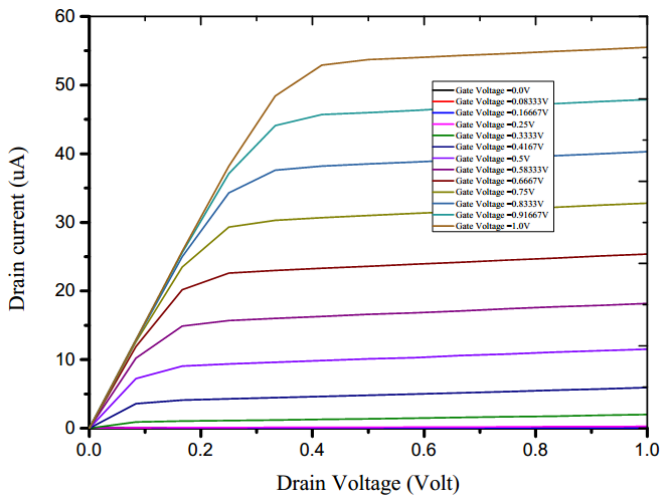


Figure 4: (b) Output Characteristic curve for a CNT having diameter 2nm and gate insulator thickness 1nm

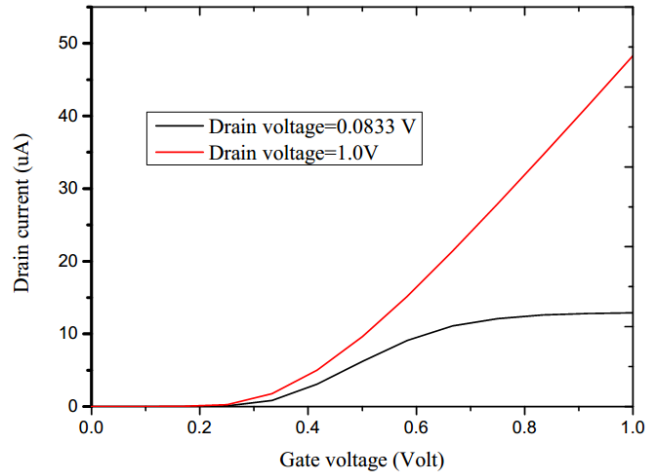


Figure 4: (c) Gate Voltage Vs Drain Current curve for a CNT having diameter 2nm and gate insulator thickness 1.5nm

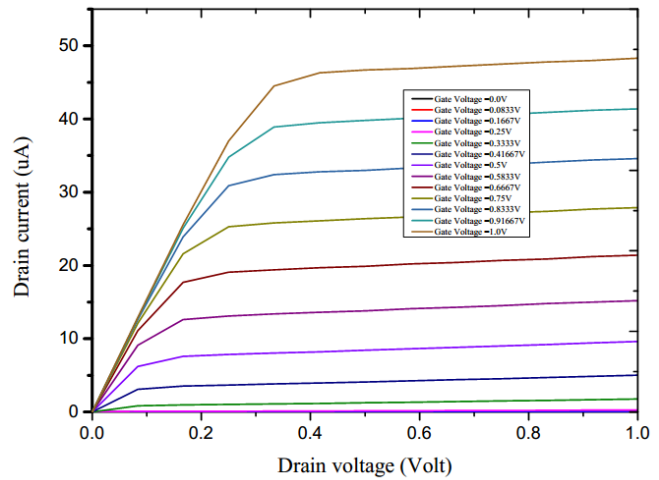


Figure 4: (d) Output Characteristic curve for a CNT having diameter 2nm and gate insulator thickness 1.5nm

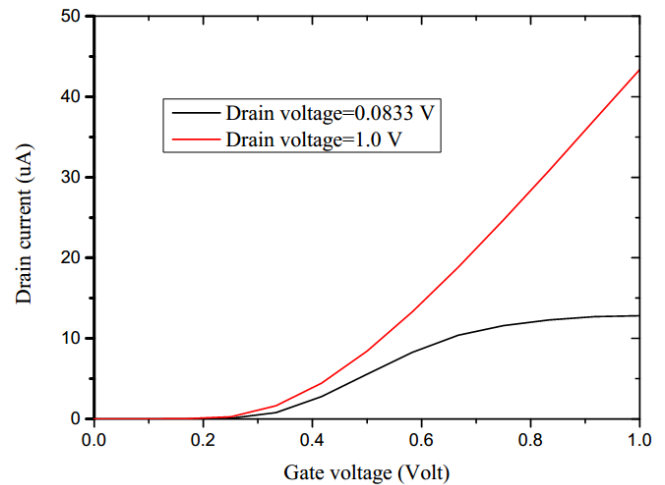


Figure 4: (e) Gate Voltage Vs Drain Current curve for a CNT having diameter 2nm and gate insulator thickness 2nm

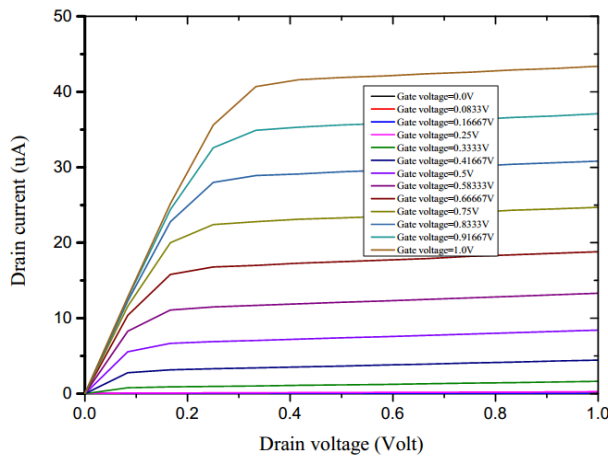


Figure 4: (f) Output Characteristic curve for a CNT having diameter 2nm and gate insulator thickness 2nm

Table 2: (Drain voltage =1 V, Nanotube diameter=2nm) [11]

Gate Voltage (V)	Drain Current (µA)		
	Gate-Insulator thickness 1nm	Gate-Insulator thickness 1.5nm	Gate-Insulator thickness 2nm
0	0.0000661	0.0000661	0.0000661
0.0833	0.00112	0.00112	0.00112
0.1667	0.0189	0.0189	0.0189
0.2500	0.28	0.271	0.265
0.3333	1.99	1.78	1.64
0.4167	5.93	5.02	4.46
0.5000	11.5	9.62	8.43
0.5833	18.2	15.2	13.3
0.6667	25.4	21.4	18.8
0.7500	32.8	27.9	24.7
0.8333	40.3	34.6	30.8
0.9167	47.9	41.4	37.1
1.0000	55.5	48.3	43.4

**4. Conclusion**

Here we can see that just by changing the nanotube diameter and gate-insulator thickness the drain current can be changed. For the nanotube diameter 1nm and gate-insulator thickness 1nm we get a huge drain current compare to the other combinations. Simulation results ensure the effect of gate dielectric increment in terms of decrement in drain current. It is due to reduction in the electric field for the same gate voltages. Thus the decay in carrier density and drain current.

This CNTFET also has a huge advantage over the Si-MOSFET. In MOSFET switching happens by altering channel resistivity whereas in CNTFET switching is due to modulation contact resistance. CNTFET generates three to four times of drive current than MOSFET. About quadruple higher transconductance of CNTFETs than MOSFETs comes from the band structure and improved mobility. The average carrier velocity is also very high almost double in CNTFET than that is in MOSFET. Its power consumption is low. Electron mobility is high. Threshold voltage

is also low. It has better control over channel formation. There is no direct tunnelling and gate leakage current is also reduced.

**Conflict of Interest**

The authors declare no conflict of interest.

**References**

- [1] Raychowdhury, A., & Roy, K. Carbon nanotube electronics: design of high-performance and low-power digital circuits. *IEEE Transactions on Circuits and Systems I: Regular Papers*, **54**(11), 2391-2401, 2007, 10.1109/TCSI.2007.907799
- [2] A. Sarkar, S. Maity, P. Chakraborty , S. K. Chakraborty , “Characterization of carbon nanotubes and its application in biomedical sensor for prostate cancer detection,” *American Scientific Publishers*, 17, 17-24, 2019, <https://doi.org/10.1166/sl.2019.4039>
- [3] A. Sarkar, M. Sreenath, K. Srinivas and NV Teja, “Investigation of Graphene as a Sensing Layer for Future Prostate Cancer Biosensing Applications,” *Journal of Physics: Conference Series*, **1921**(1), 2021, 10.1088/1742-6596/1921/1/012038
- [4] T. W. Odom, J. L. Huang et al., “Atomic structure and electronic properties of single-walled Carbon Nanotubes,” *Nature*, **391**(6662), 62-64, 1998, <https://doi.org/10.1038/34145>
- [5] S. J. Tans, A. R. M. Verschueren and C. Dekker, “Room-temperature transistor based on a single carbon nanotube,” *Nature* , **393**(6680), 49-52, 1998, <https://doi.org/10.1038/29954>
- [6] S. Rasmita and R. R. Mishra, “Simulations of carbon nanotube field effect transistors”, *International Journal of Electronic Engineering Research* **1**(2), 117-125, 2009.
- [7] N. T. Rouf, A. H. Deep and R. B. Hassan, Current-voltage characteristics of carbon nanotube field effect transistor considering non-ballistic conduction, BRAC University Dhaka-1212, Bangladesh.
- [8] S. J. Wind, J. Appenzeller, P. Avouris, “Lateral scaling in carbon nanotube field effect transistors”, *P Physical Review Letters*, **91**(5), 058301, 2003, <https://doi.org/10.1103/PhysRevLett.91.05830>
- [9] T. Dang, L. Anghel, R. Leveugle, “CNTFET Basics and Simulation. International Conference on Design and Test of Integrated Systems in Nanoscale Technology,” *IEEE Explore* 2006.
- [10] J. Guo, S. Datta and M. Lundstrom, “Assessment of silicon mos and carbon nanotube fet performance limits using a general theory of ballistic transistors,” *IEEE* 2015.
- [11] A. Rahman, J. Wang, J. Guo, Md. S. Hasan, Y. Liu, A. Matsudaira, S. S. Ahmed, S. Datta, M. Lundstrom, *FETToy*, 10254/nanohub-r220.4, 2006.
- [12] M. Radosavljevic, S. Heinze, J. Tersoff, and P. Avouris, "Drain voltage scaling in carbon nanotube transistors", *Applied Physics Letters*, **83**(12), 2435-2437., 2003, <https://doi.org/10.1063/1.1610791>
- [13] M. Zhang, P. C. H. Chan, Y. Chai, “Novel Local Silicon-Gate Carbon Nanotube Transistors Combining Silicon-on-Insulator Technology for Integration,” *IEEE transactions on nanotechnology*, **8**(2), 260-268,2009, 10.1109/TNANO.2008.2011773
- [14] Compagno, R. ed. “Technology Roadmap for Nanoelectronics,” *Microelectronics Advanced Research Initiative*, 2000.
- [15] P. Sagar P., Handa A., Kumar G., Gupta V. Nanocomposite hydrogel materials for defective cartilage repair and its mechanical tribological behavior, *A Review. Paper and Biomaterials*, **7**(3), 63-72, 2022, <https://doi.org/10.1213/j.issn.2096-2355.2022.03.007>
- [16] P. Sagar P., Handa A, Kumar G. (2022), Metallurgical, mechanical and tribological behavior of Reinforced magnesium-based composite developed Via Friction stir processing, *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*, **236**(4), 1440-1451, <https://doi.org/10.1177/09544089211063099>



## Design of an Open Source Anthropomorphic Robotic Hand for Telepresence Robot

Jittaboon Trichada, Traithep Wimonrut, Narongsak Tirasuntarakul, Eakkachai Pengwang\*

*Institute of Field Robotics, King Mongkut's University of Technology Thonburi, Bangkok, 10140, Thailand*

### ARTICLE INFO

*Article history:*

*Received: 28 September, 2022*

*Accepted: 18 December, 2022*

*Online: 24 January, 2023*

*Keywords:*

*Open source*

*Anthropomorphic Robotic Hand*

*Low cost*

### ABSTRACT

*Most anthropomorphic robotic hands use a lot of actuators to imitate the number of joints and the movement of the human hand. As a result, the forearm of the robot hand has a large size for the installation of all actuators. This robot hand is designed to reduce the number of actuators, but also retain the number of movable joints like a human hand by using the four-bar linkage mechanism and only flexion-extension movements. This stamen is added in the problem statement according to the reviewer's comment. The special features of this robotic hand are the ability to adjust the link length and the range of rotation for each joint to suit various applications and can fabricate with 3D printing and standard parts with costing about \$750. All hardware CAD files and equations are published on the GitHub website, which benefits for researchers to utilize as an open-source approach that their project might be further expanded in the future. The anthropomorphic robotic hand has five fingers, 16 joints, and 12 active Degrees of Freedom (DOFs) with 12 servo motors applied to finger motion and one for wrist motion. The structure of the hand is designed using the average of Asian human hands in combination with the golden ratio. All servo motors are installed in the forearm designed in a ventilated structure with 12V vent exhaust fan motor to stabilize the operating temperature of the robotic hand. Size and weight of the hand included with the forearm are 20×54×16.5 centimeters and 2.2 kilograms respectively. The hand has achieved human-like movement by using a four-bar linkage mechanism and tendon with PTFE tube to guide operation path of the tendon with the lowest friction force. This paper presents the design processes, the experimental set-up, and the evaluation of the finger movements. From the experiment of grasping objects, this hand was able to grasp 10 basic grasp types including 32 different objects, perform 9 common gestures, and lift the object to 450 grams. From this paper, the kinematic equation is proved that the designed finger structure can move exactly as the equation with maximum error of repeatability test around 1.6 degrees.*

### 1. Introduction

This paper is an extended paper of our work initially presented in 2021 4th International Conference on Robot Systems (ICRSA 2021) [1]. The technology that the operator interacts with the robot remotely is called the telepresence robot. This technology allows the human remotely to control the robotic end effector system in a human environment with experience as they locate there. In order to elaborate the sensibility of the operator, the design should be similar in structure and scale to the human. One of the main systems of the telepresence robot is an anthropomorphic robotic hand enabling a user to interact with the remote environment realistically.

The human hand is one of the best grippers in the world which can interact with and perceive the physical environment. However, the anatomy of the human hand is complicated to be implemented in robotics field due to size, proportions and mechanisms. For example, each person has a different size and length of each finger according to their genes [2]. The focus of this paper lies in the average hand size of Asian people, which is similar to Thai people. Thus, this paper using the average hand size of Koreans (167 males) and the Golden ratio [3-5] to design robotic hands.

Generally, the robotic hand is numerous in the design in terms of the number-type of actuator, active hand DOFs, total hand joints, power transmission, sensors, and price. The price of each robot hand varies. The cost of robotic hands has ranged between \$1,500 and \$150,000, depends on the number and type of actuator, sensor, and application programming interface (API) [6]. In

\*Corresponding Author: E. Pengwang, KMUTT, Bangkok, 10140, Thailand, (+66)24709339 eakkachai@fibo.kmutt.ac.th

addition, most robotic hands measure only grasping and performing various gestures, but this paper has added the repeatability experiments of robotic hands and the accuracy verification of the equations provided in the conference paper.

In this article, we designed an anthropomorphic robotic hand based on the anatomical structure that use the kinematic equation to calculate the length of the four-bar linkage ( $L_4$ ) of four common fingers. Before that, the designer must define basics of the configuration, including the range of motion of each joint ( $\theta_1, \theta_2, \theta_3$ ), and initial degrees ( $\alpha_2, \alpha_3$ ) as shown in Table 1 and Figure 1. Since the average Korean hand size informed the total length of each finger, a golden ratio is required to divided the average length to each phalanx ( $l_1, l_2, l_3$ ). The purpose of this paper is to design a new open-source robotic hand with a low cost (\$750) by using standard parts and 3D printing techniques and publish it on the GitHub website. Moreover, the performances of the design of low-cost robot hand are high performance from the tested as shown in the experiment section. The performances of this hand are proven from five experiments: grasping experiment, gesture experiment, motor temperature experiment, structure experiment, and repeatability experiment.

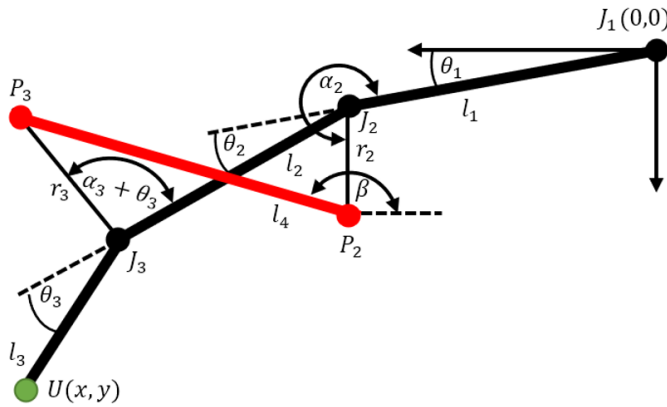


Figure 1: The Simplified Image of Finger and Parameters

Table 1: Parameter of Equations

$J_1$	Proximal joint	$\theta_1$	ROM of $J_1$
$J_2$	Middle joint	$\theta_2$	ROM of $J_2$
$J_3$	Distal joint	$\theta_3$	ROM of $J_3$
$L_1$	Proximal phalanx	$r_2$	Distance of $P_2J_2$
$L_2$	Middle phalanx	$r_3$	Distance of $P_3J_3$
$L_3$	Distal phalanx	$\alpha_2$	Initial degree of $r_2$
$L_4$	Inner link1	$\alpha_3$	Initial degree of $r_3$

## 2. Design

In fact, each joint of the human hand can move in two directions: flexion-extension and abduction-adduction. The flexion-extension movement is increase or decrease the angle between the bones of the limb at a joint while the abduction-adduction motion is away or toward the midline of the body as shown in Figure 2.

In this article, the anthropomorphic robotic right hand is 200 mm wide, 225 mm long and the forearm is 145 mm wide, 315 mm long. The robot hand contains four common fingers (differs in length of each joint), thumb, palm, wrist, forearm, and skin of

fingertips as shown in Figure 3. All fingers were designed for only flexion-extension movement. The wrist part has only one actuated joint, the thumb has three actuated joints, and the other fingers have two actuated joints per finger. Figure 2 depicts the MCP, PIP, and DIP joints on each finger, excluding the thumb. All actuated joints use a cable with PTFE tube to control a position, and the underactuated joint uses a linkage mechanism to drive the DIP joint movement that relates to the PIP joint. All servo motors are installed in the forearm designed in a ventilated structure with a 12V 4000rpm exhaust fan. The total weight is about 2.2 kilograms.

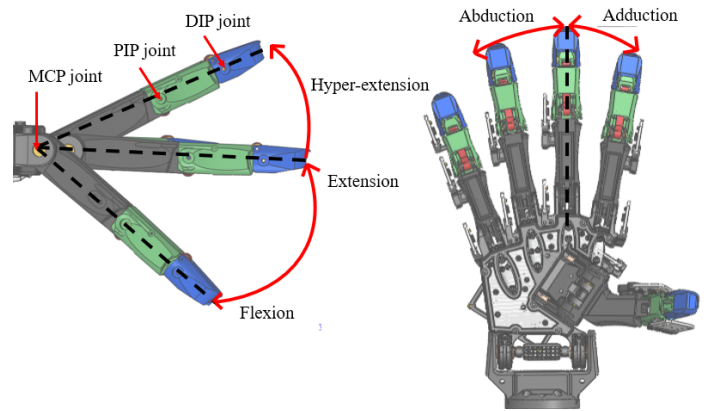


Figure 2: The Finger Movement

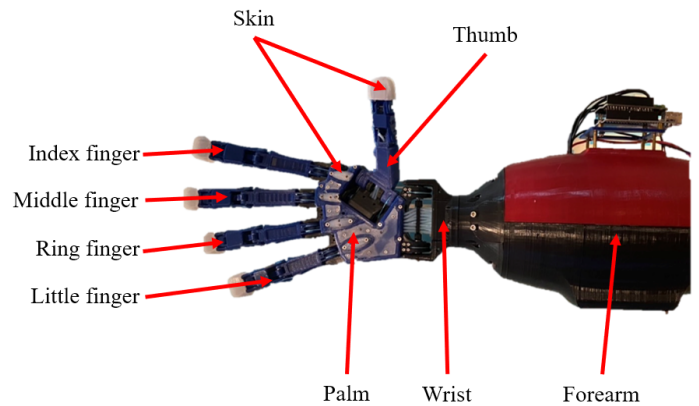


Figure 3: The Robot Hand with Forearm and Skin

### 2.1. Finger Design

This finger is a development from the finger published in the conference ICRSA 2021. Previously, tendons were routed inside robot fingers to protect tendons from the external environment, such as objects that are caught or touched that can damage the tendon. From Figure 4 (a), the distance and degree of pulling the tendons are not constant when compared with Figure 4 (b).

Because the inner finger has a very tiny space, the radius of the tendon for pulling the joints of the finger was less. To tighten the grip, the tendons used to transmit the force are shifted to the outside near the screw attachment to keep the degrees and pull distance of tendons constant throughout the operation, as shown in Figure 4. Inside the finger, there is a cavity for routing the tendon to the internal finger before attaching the PIP joint of the finger.

Since each joint has a shaft, tendons are unable to pass directly through the MCP joint. This design is enabling the tendons to pass as close as possible to the MCP joint to have the least moment of force caused by the pulling of PIP joint, and PTFE tubes can be inserted to reduce the friction force in pulling the tendons as shown in Figure 5.

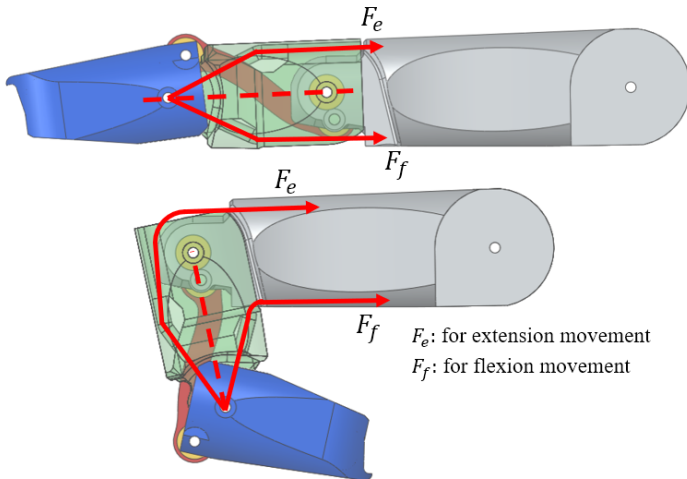


Figure 4 (a): The Index Finger and Routing Tendon Inside the Finger

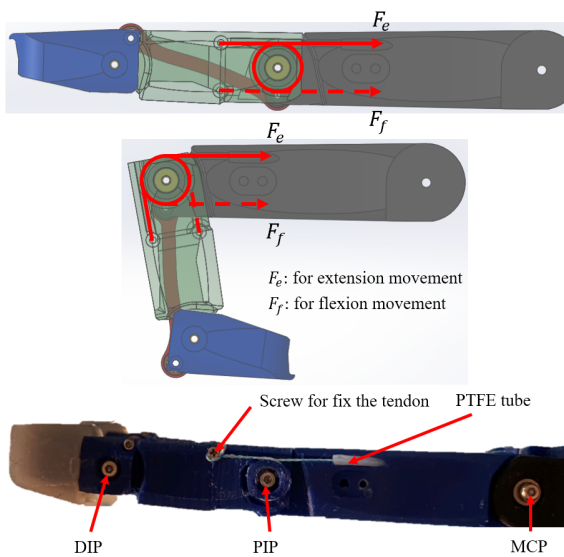


Figure 4 (b): The Index Finger with Tendon Routing Outside and Fingertip's Skin

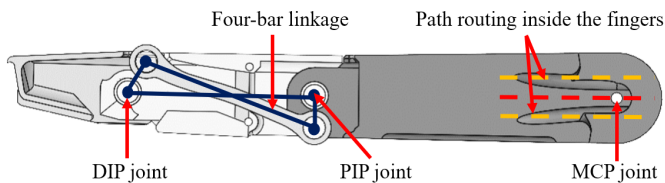


Figure 5: The Four-bar Linkage Mechanism and Path Routing in the Finger

Because the average size of the Korean hand only defines the length of each finger, the golden ratio is required to calculate the length of each phalanx. The distal phalanx of each finger is very

small of different lengths (21.15 mm, 22.01 mm, 21.08 mm, 18.45 mm), making it unable to be installed with a four-bar linkage mechanism. This problem can be solved by using an equal length of each distal phalanx. The length of distal phalanx that can be achieved is 21.37 mm. Each finger's length and range of motion are uniformly designed, as shown in Tables 2, 3, and Figure 6 respectively.

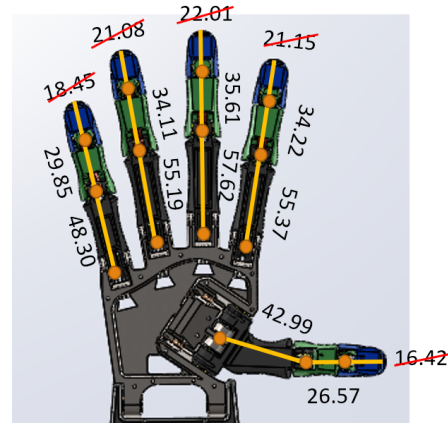


Figure 6: The Length of Each Phalanx

Table 2: Length of Each Phalanx [mm.]

No.	Name	Length (mm.)
1	Distal phalanx-all fingers	21.37
2	Middle Phalanx-Index	34.22
3	Middle Phalanx-Middle	35.61
4	Middle Phalanx-Ring	34.11
5	Middle Phalanx-little	29.85
6	Proximal Phalanx-Thumb	26.57
7	Proximal Phalanx-Index	55.37
8	Proximal Phalanx-Middle	57.62
9	Proximal Phalanx-Ring	55.19
10	Proximal Phalanx-little	48.30
11	Metacarpal Phalanx-Thumb	42.99
12	Distal phalanx wide-all fingers	18.00
13	Distal phalanx thin-all fingers	13.00
14	Middle phalanx wide-all fingers	20.50
15	Middle phalanx thin-all fingers	14.50
16	Proximal phalanx wide-all fingers	20.50
17	Proximal phalanx thin-all fingers	16.00
18	Metacarpal phalanx wide-all fingers	23.50
19	Metacarpal phalanx thin-all fingers	15.50

Table 3: The Range of Motion of Each Joint [degree]

Name	DIP joint	PIP joint	MCP joint	DIP-T joint	MCP-T joint	CMC-T joint
Range of Motion	0 to 80	0 to 90	-10 to 90	0 to 100	0 to 100	-10 to 120

The human hand contains at least 23 degrees of freedom (DOFs) [7]. Human fingers have 3 joints with 4 DOFs: 3 DOFs for flexion-extension movement and 1 DOF for adduction-abduction

movement. In the case of the thumb, there are 3 joints with 5 DOFs: 3 DOFs for flexion-extension movement and 2 DOFs for adduction-abduction movement (Figure 7). The wrist has 2 DOFs for flexing and expanding (Figure 8). The robotic hand has five fingers, 16 joints, and 12 active DOFs (only flexion-extension movement) with 12 servo motors. Four common fingers excluding the thumb use 2 servo motors per finger to control the PIP joint and MCP joints, while the DIP joint moves related to the PIP joint by using the four-bar linkage mechanism as shown in Figure 5. The thumb uses 3 servo motors and 1 servo motor for the wrist. Each movable joint use bearing to reduce the friction force during finger movement. The MCP joint of 4 fingers and the CMC joint of thumb use 2 mm inner diameter bearing. Others moving joints in each finger use 1.5 mm inner diameter bearing. The length of the four-bar linkage of the four fingers was calculated from equations  $r_3$  and  $l_4$  in the conference as shown in Table 4. Each four-bar linkage structure has a different concave curvature to prevent the linkage from a collision with the shaft of each joint while operating (Figure 9).

Table 4: Length of Each Four-bar Linkage [mm.]

Name	Length (mm)
Linkage Index	33.04
Linkage Middle	34.34
Linkage Ring	32.93
Linkage Little	28.98

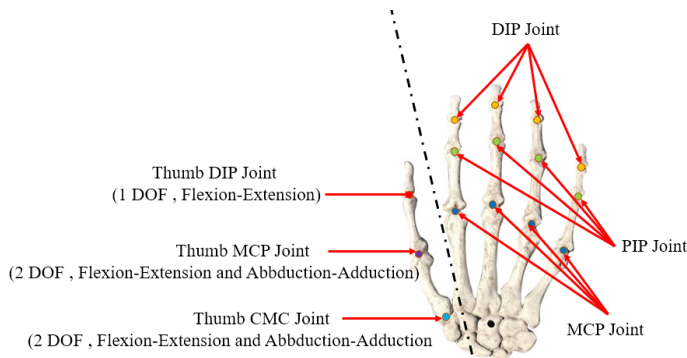


Figure 7: The Anatomy of Human Hand

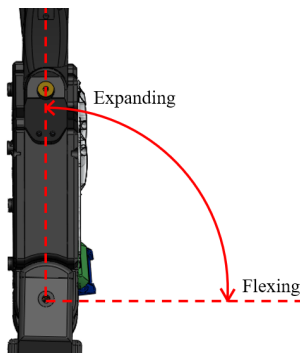


Figure 8: The Wrist Movement

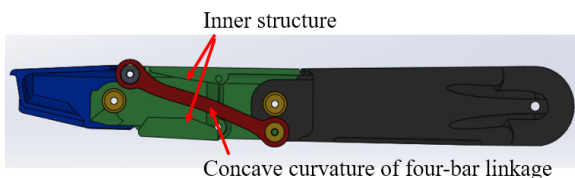
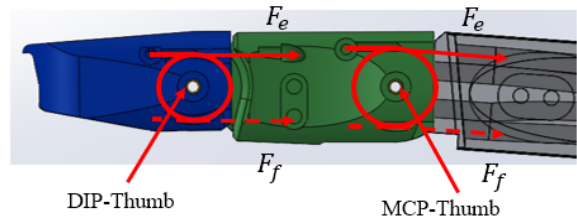


Figure 9: The Curvature of Four-bar Linkage

Many robotic hands have various thumb designs to accommodate the suitability of the hand, such as the number of DOFs, actuators, power transmission, and functionalities; thus, the design of the thumb is rarely defined. The main contribution for this paper is separated into 2 sections: the joint design, and the structure design. The joint design of the thumb is the same design as the controllable joint of four fingers which has a constant both distance and degree for pulling as shown in Figure 10. Normally, the thumb grasps an object by using both abduction-adduction movement and flexion-extension movement. The flexion-extension movement of thumb is used to press objects towards the palm of the hand while the abduction-adduction movement of thumb is used to grasp the various objects. Even though the abduction-adduction motion is very important for grasp [8], this robot hand challenges to design by using only flexion-extension movement to achieve the purpose of this hand in order to reduce the number of motors to be installed on forearm. From the reduction to 1 DOF per joint of the thumb, the importance of thumb angles must be emphasized, which affects the grasping performance of the hand. The motion analysis in SolidWorks program is required to test basic gestures and basic grasps such as handfuls, index-thumb, middle-thumb, cylindrical grasp, and spherical grasp before forming. From the analysis, the best angle for structure design of the thumb is 58.5 degrees in the top view and 28.5 degrees in the front view (Figure 11) to perform as many different gestures and basic grasps as possible in motion analysis of the SolidWorks program.



$F_e$ : for extension movement  
 $F_f$ : for flexion movement

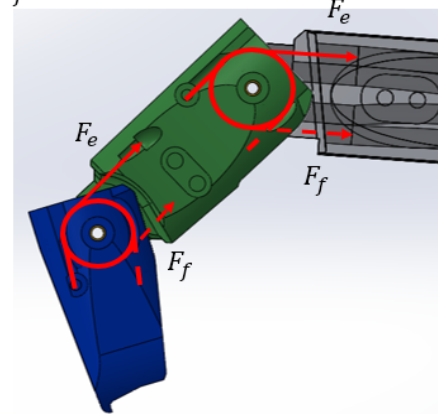


Figure 10: The Joint Design of the Thumb

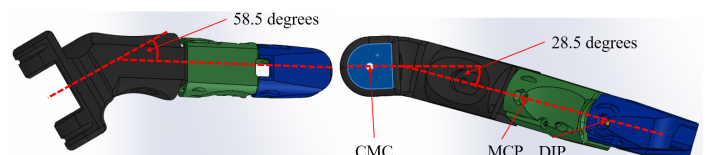


Figure 11: The Attachment Angle of the Thumb

## 2.2. Palm and Wrist Design

Generally, human hands can grasp objects by using the palm, all fingers, wrist, and skin. The palm is the part that connects the four fingers, thumb, and wrist that is coated by human skin. In the robotic field, the robot finger with no adduction-abduction movement was fixed angle between neighbor MCP joint of each finger at about 12 degrees [9]. The angle of neighbor MCP joint allows the hand to tightly grasp objects, increasing the area of routing tendons inside the hand, and decreasing the bending of PTFE tube around the MCP joint of the index finger in the palm. Each robotic hand has a different angle between neighbor MCP joint depending on the hand. From the analysis, it was found that 10 degrees is most suitable for this hand to gesture and grasp. The palm is designed by defining the middle finger as the reference point with 10 degrees of angle between neighbor MCP joints as shown in Figure 12.

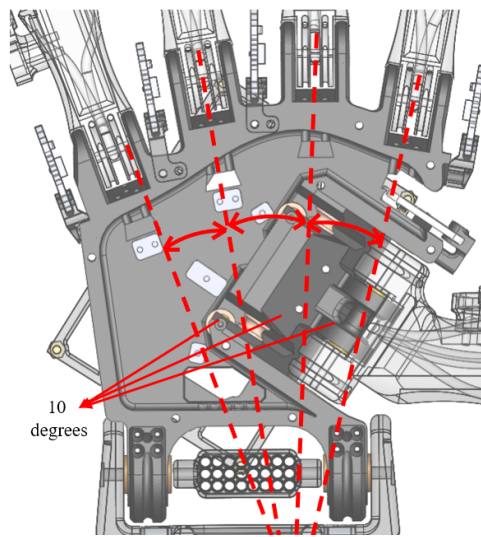


Figure 12: The Palm

In the palm, there is an apparatus that is used to arrange the PTFE tube and tendons from the wrist to each finger as shown in Figure 13. The CMC joint of the thumb was designed to angle the middle finger about 30 degrees (Figure 13) so that PTFE tubes inside the palm are easy to assemble and do not bend too much. The OK pose and check gesture can be performed by using this design. The OK posture is the index finger and thumb converge while the others are fully spread as Figure 14. The check posture is the index finger and thumb are fully spread while the others are fully bent same as the check symbol as shown in Figure 15. These 2 poses are the basic postures to hold objects and make various gestures of the hands. Since the direction of rotation of the CMC joint of the thumb is in a different direction of the tendon movement to other fingers, The bearing must be provided for changing the direction of pulling the tendons of the thumb as shown in Figure 16.

The kit for re-arranging the PTFE tube with tendon

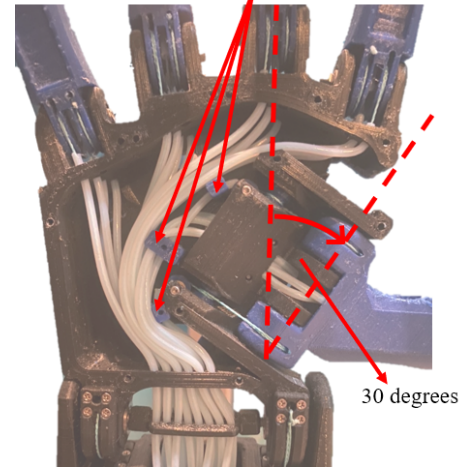


Figure 13: The Re-arranging of the PTFE Tube in the Palm

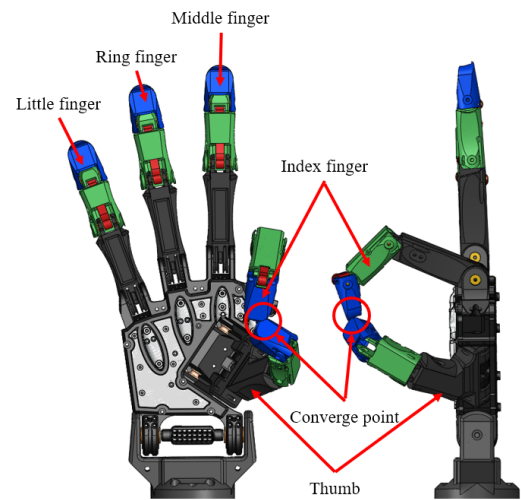


Figure 14: The Robotic Hand Performs OK Gesture

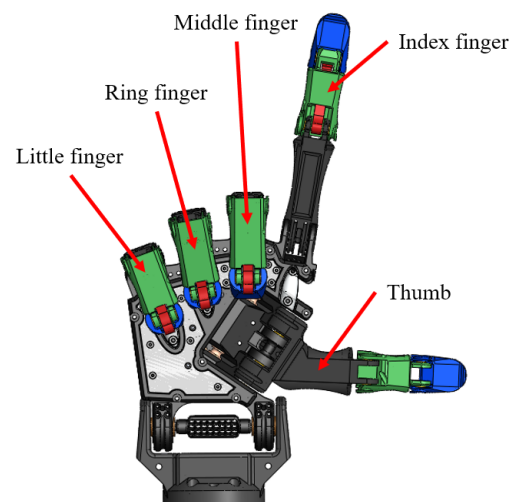


Figure 15: The Robotic Hand Performs Check Gesture

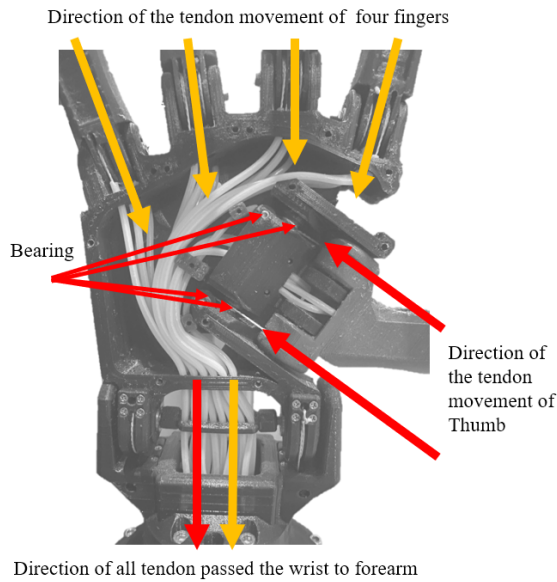


Figure 16: Direction of Tendon Movement in the Palm

PCB standoff spacers and screw (Figure 18). The palm is assembled with an extension palm part on the wrist area to increase radius for pulling the tendon (Figure 19).

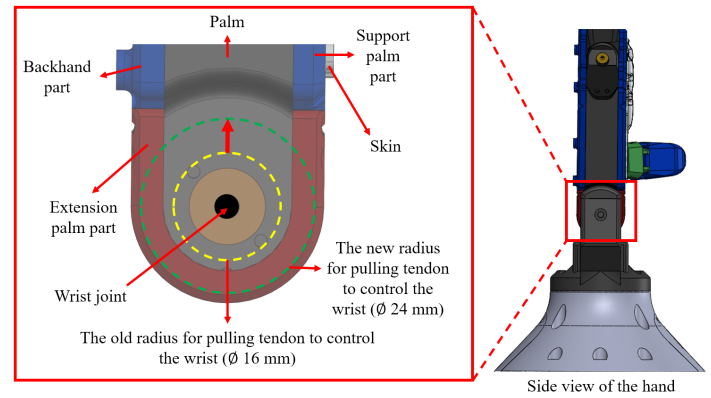


Figure 19: The Extension Palm Part

The wrist area has 3 components which are the main wrist, extension palm, and a support PTFE tube kit as shown in Figure 20. The main wrist is designed by integrating with PTFE tube to decrease friction force for pulling the tendon of wrist joint. In the middle of wrist joint is a support PTFE tube kit for re-arranging and supporting the PTFE tubes with tendons when the tendon is pulled by the motor. This support PTFE tube kit will keep the PTFE tube in place no matter how many degrees the wrist is tilted. The motor that controls the wrist joint will reduce the load caused by the tendon movement of all fingers. Main wrist is a connector between the hand and forearm that is used to re-arranging the PTFE tube same as a support PTFE tube kit and guides the tendon to attach the wrist joint. The range of motion of the wrist joint is from 0 to 180 degrees, the same as the wrist joint of human hand as shown in Figure 21.

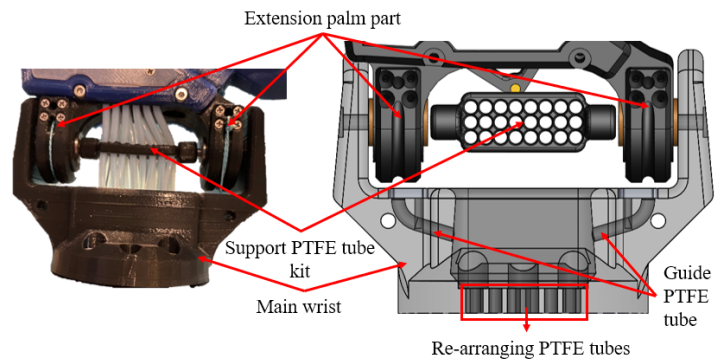


Figure 20: The Wrist

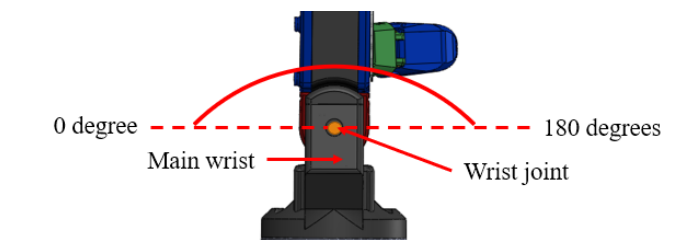


Figure 21: The Range of Motion of Wrist Joint

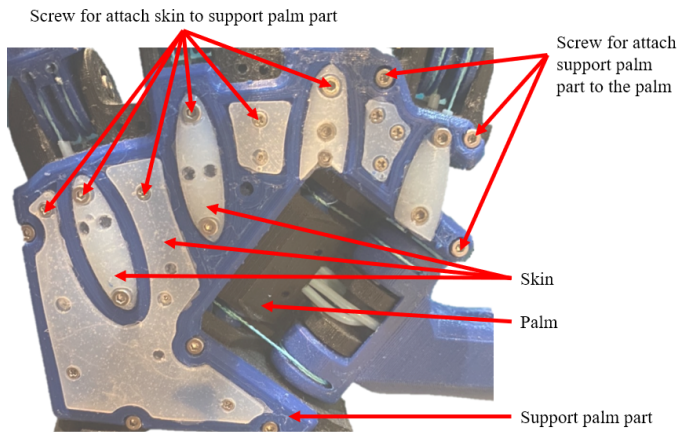


Figure 17: The Support Palm Part

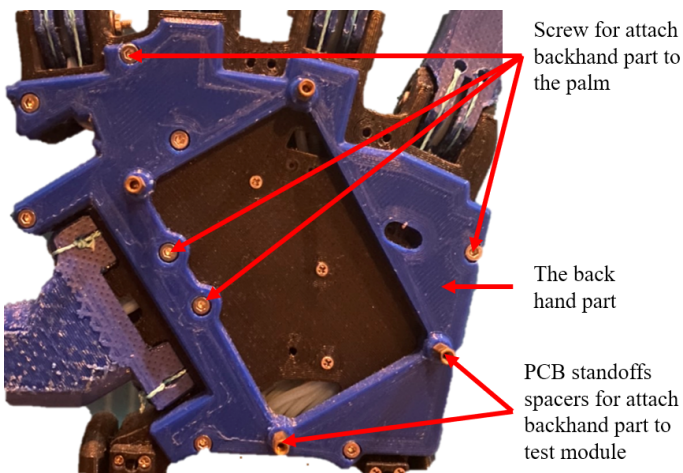


Figure 18: The Back Hand

The support palm part is the connector between palm and skin by using the screw (Figure 17). The backhand part attached behind the palm is used to install the test module (PCB for the connection wire of the encoder, multiplexer, and Arduino Uno board) by using

### 2.3. Skin of Fingertips and Palm

The entire skin is made of silicone (RA-22AB, RUNGART, Thailand) [10] forming with the use of PLA moles (3D printing). The skin of the palm consists of flat palm skin and half-ellipse palm skin (Figure 22). The thickness of flat palm skins and half-ellipse palm skins are about 1.5 mm and 4 mm respectively as shown in Figure 23. The skin of fingertips is the one of the main parts for grasping an object. Skin tips are used to increase friction force for tightening grip. These skin tips are wearable skin that has the same shape as fingertips with a thickness from the outside of the fingertips of about 2.5 mm as shown in Figure 24.

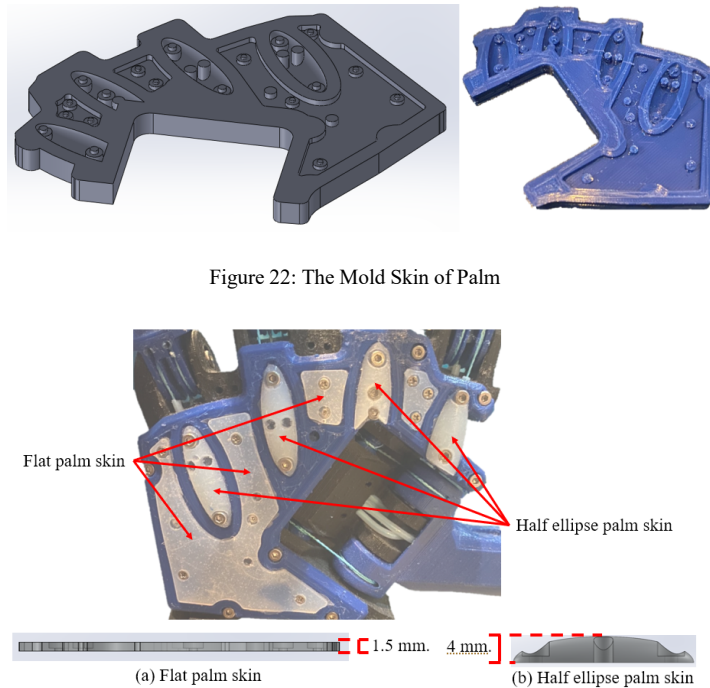


Figure 22: The Mold Skin of Palm

Figure 23: The Thickness of Palm Skin

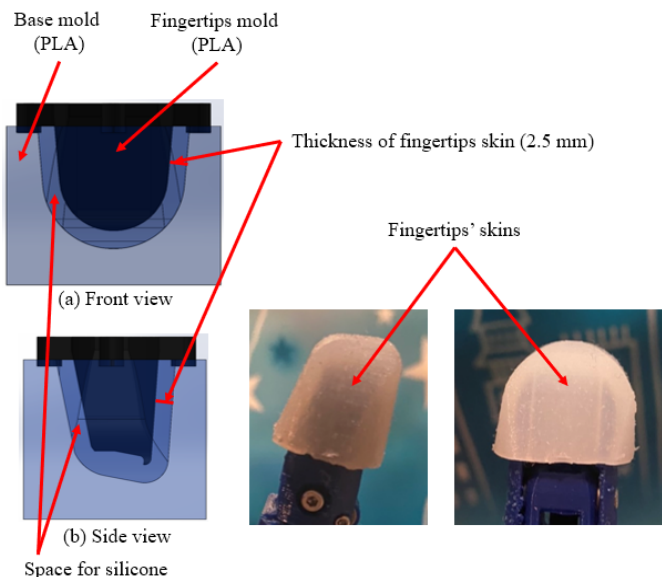


Figure 24: The Mold Skin of Fingertips

### 2.4. Forearm Design

The forearm contains the main forearm, connecting motor part, PTFE tube guide part, and cover. All servo motors, Arduino Uno board with Dynamixel shield, 12V fan, and PCB boards are installed in the forearm as shown in Figure 25. This Dynamixel servo motor (XL430 W250-t) has engineering plastic gear with an operating temperature from  $-5$  to  $+72$  °C. Long periods of heavy work of the motor will cause the accumulation of high temperature inside the forearm, so the design of the robot forearm must focus on temperature reduction to extend the motor life. The structure of the forearm must be a ventilated structure with a 12V vent exhaust fan to stabilize the operating temperature of the robotic hand. All PTFE tubes with tendons are routed following the PTFE guide from the motor in the forearm to each joint of the finger and wrist.

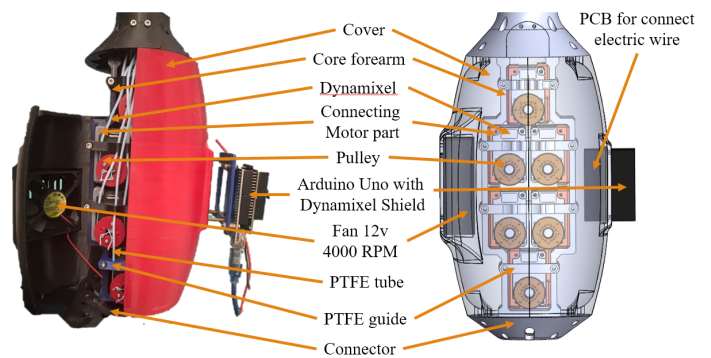


Figure 25: The Forearm

### 2.5. Actuation, Power Transmission and PLA Material

This hand uses 12 servo motors to control each actuated joint while underactuated joints (DIP joint of four fingers) move to follow the actuated joints (PIP joint of four fingers) by using a four-bar linkage mechanism. The actuators can transmit the power to control each joint by using a tendon with the PTFE tube. The PTFE tube [11] is used to protect the tendon and reduce the friction force between the tendon and the body part (PLA). The actuators of this robotic hand are XL430 W250-T from Dynamixel (TTL connection) because it has a suitable price with high torque and small size (Table 5). In addition, this actuator has precision to control with feedback sensors: position, current load, current temperature, etc. The tendon that is used to transmit the power from the actuators is a fishing tendon from Proberos. This tendon is made from 4 braids of Polyethylene (PE) with high max tension (36.2 Kg) and a small diameter (0.5 mm) costing about \$1.5 per 100m as shown in Table 6. Because PLA filament has Young's modulus of about 3.04 GPa but ABS material has Young's modulus of about 1.97 GPa [12]. Therefore, the material for 3D printing parts uses PLA (Polylactic acid) filament with a cost (of \$17.99 per Kg.) [13-14].

Table 5: Specification of Servo Motor [15]

Name	Dynamixel (XL430 W250-T)
Stall Torque	1.5 N.m
Stall Current	1.4 A
Weight	57 grams
Dimension	28.5 x 46.5 x 34 mm.

No Load Speed (at 12V)	61 rev/min
Resolution	4096 pulse/rev, 360 degree
Gear Ratio	258.5 : 1
Operating Temperature	-5 to +72 °C
Connection	TTL
Feedback	Position, Load, Temperature, etc.
Material	Engineering Plastic
Price	\$49.90

Table 6: Specification of Tendon [16]

Name	Tendon
Brand	Proberos
Material	Polyethylene (PE)
Number of tendons	4 braids
Outer Diameter	0.5 mm
Max Tension	36.2 Kg.
Price (100 m.)	\$1.5

### 2.6. Cost Analysis

Generally, the range cost of a robotic hand is between \$1,500 and \$150,000. However, this robotic hand costs about \$750 (as shown in Table 7) with 12 servo motors (Dynamixel) that can grasp objects and gestures similar to robotic hands that cost \$1500 as proof in the results and experiment section. All parts of this robot hand are made from 3d print, designed for standard components that can be easily purchased locally and replaced. Moreover, robotic hands are designed to use as few motors as possible while keeping the ability to grasp various objects and gestures like other robotic hands as much as possible. The four-bar linkage mechanism is used to control the DIP joint related to the PIP joint in four fingers that can reduce one actuator per finger with the same number of movable joints as shown in the design section. Because all actuators are installed in the forearm, the low number of actuators in use saves costs and reduces the size and weight of the forearm. Thus, the price of this hand will be cheaper than the others.

Table 7: Price of Actuator and Material

Name	Amount	Total Price (\$)
Dynamixel XL430 W250-T	12	598.80
PLA (eSUN) 1 Kg.	2	35.98
Arduino Uno	1	20
Dynamixel shield	1	19
Fan 12V 4000 RPM	1	1.5
Power supply 12V 20A 240W	1	9
LCD meter and shunt (20A)	1	9
Electric wire AWG24 (30m.)	2	5.50
Tendon	1	1.5
Bolt and screw	-	15
Bearing, etc.	-	35
Total		750.28

### 3. Experiment and Results

All experiments are intended to prove the various performances of robotic hands compared to other expensive robotic hands such as grasping objects and gestures. The development of the anthropomorphic robotic hand in this paper has [www.astesj.com](http://www.astesj.com)

five experiments: grasping experiment, gesture experiment, motor temperature experiment, structure experiment, and repeatability experiment.

One of the performance experiments of the anthropomorphic robotic hand is the grasping experiment that uses various objects in daily life to test the grasping of the robot hand. The robotic hand grasp objects that are different in shape, weight, and size by using various grasping gestures as shown in the result of the grasping experiment.

The gesture experiment is the test of the robot hand to perform basic hand gestures and symbols that were chosen from daily hand posture. These two experiments were intended to test the ability to grip objects and perform gestures as designed.

The third experiment is the operating motor temperature test to determine whether the added structure and fan can reduce the temperature of the motor while operating. The motor temperature experiment uses Arduino Uno to read the current temperature from the feedback sensor of each motor.

The fourth experiment is the structure test of the degrees of dip and pip joints of the index finger, whether the DIP joint moves along with the PIP joint by using the four-bar linkage mechanism is similar to the equation used in the design.

The last experiment is the repeatability test of the robotic hand which shows how many errors each joint has. The structure experiment and repeatability experiment use magnetic encoders (AS5600) and an Arduino Uno board to read the current degree of each joint.

All experiments test only the joints of the fingers excluding the wrist. The controllable joints of this robot hand are 11 joints (3 joints for the thumb and 2 joints for each finger), therefore the maximum magnetic encoder used to read the angle of each joint is 11 positions. The Arduino Uno board connects to each magnetic encoder board by using I2C communication. All encoders cannot connect to the Arduino Uno board directly because each encoder has the same address (0x36). The I2C multiplexer (TCA9548A) is required to expand the I2C bus port and control multiple I2C devices with the same I2C address. One multiplexer can connect to 8 devices, so 2 multiplexers are enough. The address of the multiplexer can select a value from 0x70 to 0x77 by adjusting the values of the A0, A1, and A2 pins. The robot hand with an encoder module for the test is shown in Figure 26.

The average error values of the structure experiment and repeatability experiment are calculated from the following equation.

$$Average\ Error\ Value = \frac{\sum_{i=0}^n (X_n - T)}{N}$$

Table 8: The Meaning of Variable

Variable	Meaning
$X_n$	Position value from encoder ( $n^{th}$ )
$T$	Target position value
$N$	Total of test



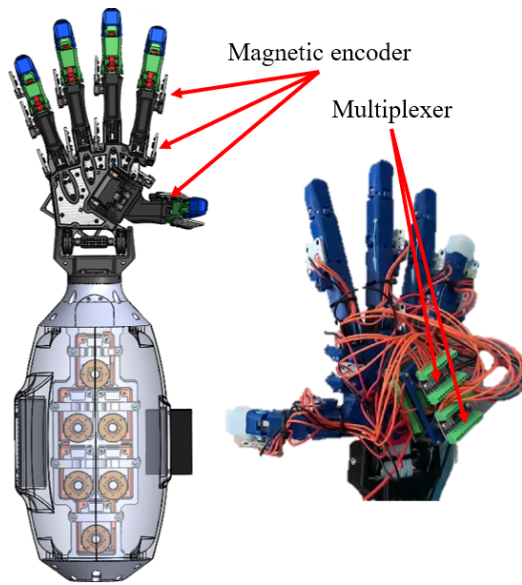


Figure 26: The Robotic Hand with Encoder Module

### 3.1. Result of Grasping Experiment

Generally, there are two kinds of grasping objects: power grasp and precision grasp. The precision grasp (tripod, two fingers, disk, and tip pinch) uses only fingertips with skin to hold small lightweight objects, while the power grasp (spherical, cylindrical, lateral pinch, lumbrical, large diameter, and platform) uses every part of the hand (fingertips, phalanx, palm, and skin) to grasp huge heavyweight objects. This experiment is the gripping test of the robot hand with power grasp and precision grasp by using ten grasping gestures: spherical (1.01-1.02), tripod (2.01-2.04), two fingers (3.01-3.02), cylindrical (4.01-4.05), lateral pinch (5.01-5.11), lumbrical (6.01), disk (7.01), large diameter (8.01-8.04), tip pinch (9.01), and platform (10.01) as shown in Tables 9 and Figure 27 respectively.

Because this robotic hand is controlled by humans to be used for handling various objects in daily life, The items used in the test must be found in daily life (differ in shape, weight, and size). There are 32 different objects that are used to test such as baseball, glue, pen, table tennis, bottle 600 ml, screwdriver, power bank, key, lighter, book, disk, and plaster. This test only focuses on that each item can be held in that posture without getting out of hand. The robot hand successfully grasped selected 32 different objects with 10 basic postures and can grasp objects up to 450 grams (bottle 600 ml) in cylindrical gripping gesture. The bottle made from plastic (PE) with a slippery skin and large diameter is grasped by cylindrical gesture (power grasp) to test the maximum weight that this hand can hold and to test whether robotic hands can handle things (not structural polishing). Holding a slippery object in this pose is a real gripping efficiency test of the robotic hand because the object may slip out of hand if the grasp is not tight enough. The proposed anthropomorphic design allows our robotic hand to grasp objects in a suitable gripping posture.

Table 9: Grasping Poses and Objects (a)

Grasping Pose	Objects		
	Name	Dimension(mm)	Weight(g)
	Baseball (1.01)	Ø 73	150

Spherical (1)	Tennis ball (1.02)	Ø 65	55
Tripod (2)	Glue (2.01)	Ø 20	11
	Pencil (2.02)	Ø 7.8	4
	Pen (2.03)	Ø 9.8	6
	Marker (2.04)	Ø 10	7

Table 9: Grasping Poses and Objects (b)

Grasping Pose	Objects		
	Name	Dimension(mm)	Weight(g)
Two Fingers (3)	Table tennis ball (3.01)	Ø 39.5	2
	Golf ball (3.02)	Ø 42.5	45
Cylindrical (4)	Bottle 600 ml (4.01)	Ø 60	450
	Bottle skin care (4.02)	Ø 50	72
	Huge screwdriver (4.03)	Ø 33.5	96
	Trowel (4.04)	Ø 32	27
	Power bank (4.05)	Ø 41.5	133
Lateral Pinch (5)	Key (5.01)	Thick 4.9	6
	Smart key (5.02)	Thick 0.8	4
	Metal key (5.03)	Thick 2.2	39
	Card reader (5.04)	Thick 8.5	3
	Tweezers (5.05)	Thick 10	15
	Small screwdriver (5.06)	Ø 7.1	14
	Pen (5.07)	Ø 9.8	27
	Hand drill (5.08)	Ø 8.15	41
	Lighter (5.09)	Thick 11	13
	Tape (5.10)	Thick 18.5	22
	Utility knife (5.11)	Ø 9	15
Lumbrical (6)	Book (6.01)	148.5 × 210 Thick 12	110
Disk (7)	Disk (7.01)	Ø 19 Thick 1.25	17
Large Diameter (8)	Wire strippers (8.01)	104 × 15	174
	Combination pliers (8.02)	90 × 16.5	200
	Diagonal cutter (8.03)	93 × 11.5	25
	Screwdriver box (8.04)	67.5 × 17.25	263
Tip Pinch (9)	Wound closure plaster (9.01)	Thick 1	1
Platform (10)	Document pouch (10.01)	297 × 210 Thick 8	250

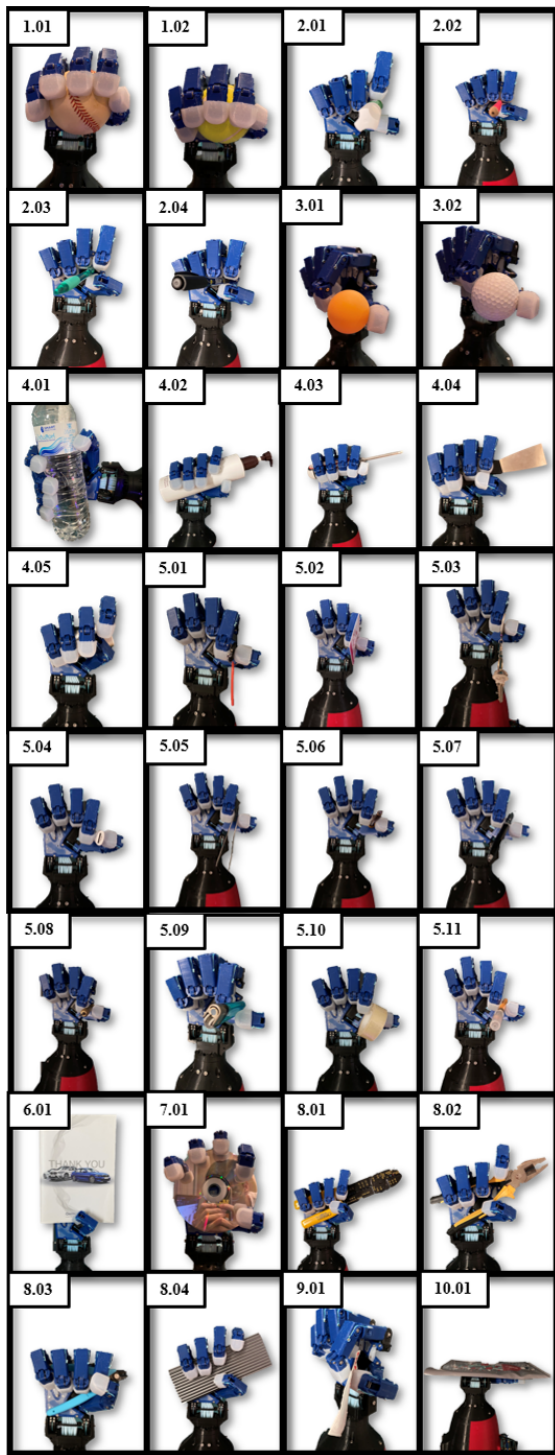


Figure 27: Robotic Hand Grasping Various Objects

### 3.2. Result of Gesture Experiment

This experiment is a test of the basic hand gestures and symbols that are chosen from frequently used in daily life. The robotic hand successfully posed 9 common gestures including high-five (1), peace (2), ok (3), index pointing (4), grasp (5), promise (6), love (7), check (8), and good job (9) as shown in Figure 28. The purpose of this robotic hand design does not focus on the adduction-abduction movement but to reduce the number of motors.

Consequently, this hand cannot perform gestures that use the abduction-adduction movement such as fingers crossed, fig sign, and Vulcan salute.

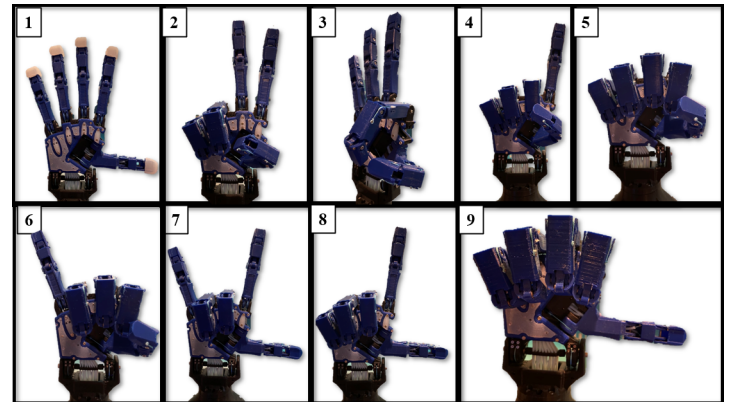


Figure 28: Gesture of Robot Hand

### 3.3. Operating Temperature of Motor Experiment

This experiment was to prove that an extra fan can reduce the temperature of the motor while operating by monitoring the motor temperature directly from the motor feedback sensor. The fan is installed to the cover of the forearm as shown in Figure 25. This is the performance experiment of a designed ventilated structure with a fan (12V 4000RPM). Usually, the motor can operate in the temperature range between -5 to +72 °C, but the operating motor temperature at 30 percent torque (enough for grasping objects) is between 55.0 °C to 68.0 °C (Figure 29) that close to the maximum temperature of the motor. This experiment uses Arduino Uno to control 12 motors and get feedback (Temperature) from motors in real-time (20 times). After installation and test, the range temperature of the operating motor is between 46.0 °C to 56.0 °C as shown in Figure 30. From the above, this experiment can prove that the fan can reduce the average temperature of the operating motor from 61.5 °C to 51.0 °C.

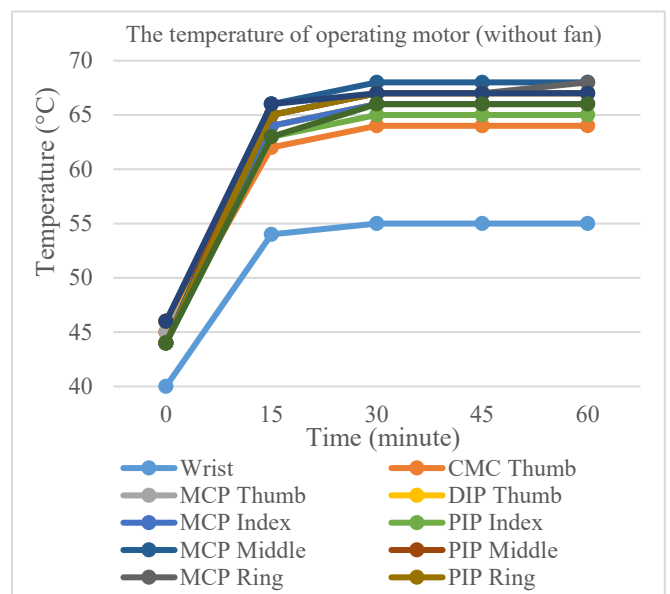


Figure 29: The Temperature of Each Operating Motor in Solid Structure

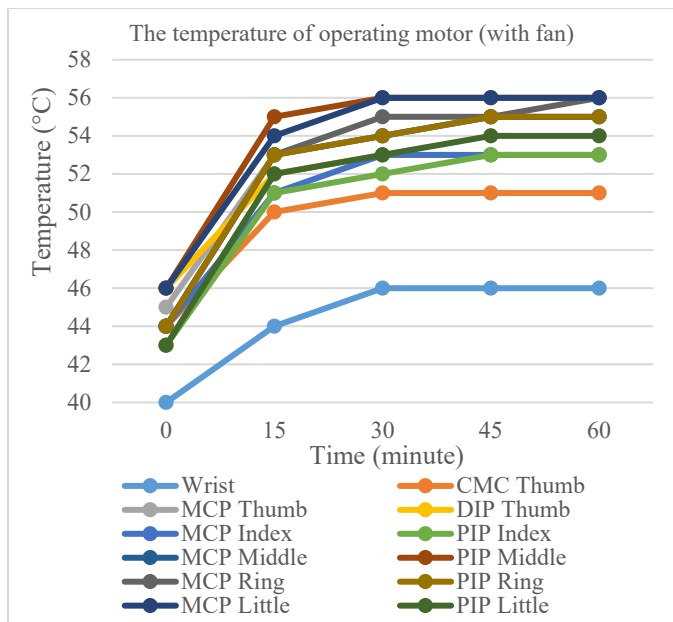


Figure 30: The Temperature of Each Operating Motor in Ventilated Structure

### 3.4. Structural of Four Common Fingers Experiment

The kinematic equation is used to design a four-bar linkage mechanism to move dip joints related to pip joints by using 1 actuator to control. This experiment is a structural experiment that tests the movement of dip joints related to pip joints (index finger) compared with the equation used in the design. This test uses magnetic encoder and Arduino Uno to check the degrees of each joint and control servo motors. The test will control the PIP joint and observe DIP joint of four common fingers around 100 times per position and compare with the kinematic equations from [1]. While the PIP joint is controlled by servo motor which moves from 0.0 degree to 78.0 degrees and back to 0.0 degree (100 times), the DIP joint moves from 0.0 degree to 68.1 degrees following the PIP joint by using a four-bar linkage mechanism. When the PIP joint moves from 0.0 degree to 78.0 degrees, the DIP joint will move from 0 to 69.3 degrees by calculating from the equation. After testing, the structure can move according to the equation with an error of fewer than 1.6 degrees and the PIP joint of index finger has an error of about 0.1 degrees as shown in Tables 10 and 11.

Table 10: Structural Test (Degree)

No	Encoder (statistic method)				Equation	Avg. Error (DIP)
	PIP Joint (Deg)		DIP Joint (Deg)		DIP Joint (Deg)	
	Average	SD	Average	SD		
1	78.1	0.2	68.1	0.2	69.3	1.2
2	31.1	0.2	29.1	0.1	27.5	1.6

Table 11: The Average Error of PIP Joint (Degree)

No	Encoder (statistic method)		Target	Avg Error (PIP)
	PIP Joint (Deg)		PIP Joint (Deg)	
	Average	SD		
1	78.1	0.2	78.0	0.1
2	31.1	0.2	31.0	0.1

### 3.5. Repeatability Experiment

The repeatability test of robotic hands using Arduino Uno boards to control 2 positions of the motor and read the position of each joint from magnetic encoder modules (12-bit resolution or about 0.08° per count). The Arduino Uno board can read the degree of each joint in real-time via a magnetic encoder. This board control motor moves forward and backward to the same position around 200 times per cycle. To conclude the data in the table, all information is expressed as the statistical method (max, min, standard deviation, average). This test is divided into two experiments which are the repeatability test of the index finger, and the repeatability test of the robot hand. From all of the experiments, this robot hand has a maximum error of repeatability of about 1.2 degrees.

First, the repeatability test of the index finger uses 3 magnetic encoders with Arduino Uno to measure the degree of MCP, PIP, and DIP joint of the index finger. This test has three sets of moving positions and each set has two positions that are selected from the range of motion of each joint. From the result, we found that the maximum error of the repeatability test of the index finger is 0.2 degrees as shown in Table 12.

Table 12: Repeatability Test of Index Finger (Degree)

Number of Sets		Statistical Method				Target	Avg Error
		Min	Max	Avg	SD		
MCP (1)	Pos 1	0.0	0.6	0.1	0.1	0.0	0.1
	Pos 2	84.0	85.1	84.1	0.2	84.0	0.1
PIP (1)	Pos 1	0.0	2.0	0.1	0.2	0.0	0.1
	Pos 2	78.0	79.8	78.2	0.4	78.0	0.2
MCP (2)	Pos 1	0.0	0.5	0.1	0.1	0.0	0.1
	Pos 2	55.0	56.1	55.1	0.2	55.0	0.1
PIP (2)	Pos 1	0.0	0.5	0.1	0.1	0.0	0.1
	Pos 2	67.0	69.0	68.2	0.3	68.0	0.2
MCP (3)	Pos 1	0.0	0.5	0.1	0.1	0.0	0.1
	Pos 2	70.0	72.0	70.2	0.4	70.0	0.2
PIP (3)	Pos 1	0.0	0.5	0.1	0.1	0.0	0.1
	Pos 2	34.8	36.2	35.1	0.2	35.0	0.1

Second, the repeatability experiment of the robot hand is testing the error of all controllable joints to find the maximum error of the repeatability test of the robot hand by using 11 magnetic encoders and Arduino Uno. This test has settings and methods the same as the repeatability test of the index finger. The maximum error of this test is 1.2 degrees at CMC joint of thumb as shown in Tables 13 and 14. The PTFE tubes with tendons are routed through the CMC joint to control the MCP and DIP joints of the thumb while the other joints have only tendons that route through. The maximum error of the repeatability test is on the CMC joint.

Table 13: Repeatability Test of Robot Hand (Degree)

Finger Names	Name of Joint		Statistical Method (Deg)			
			Min	Max	Avg.	SD
Thumb	CMC	Pos 1	0.0	2.0	0.1	0.2
		Pos 2	98.0	100.0	99.2	0.3
	MCP	Pos 1	0.0	0.5	0.1	0.1
		Pos 2	76.0	77.0	76.1	0.1
	DIP	Pos 1	0.0	2.0	0.1	0.2

Index	MCP	Pos 2	39.0	40.0	39.1	0.2
		Pos 1	0.0	0.5	0.1	0.1
	Pos 2	70.0	71.3	70.3	0.5	
Middle	MCP	Pos 1	0.0	0.6	0.1	0.1
		Pos 2	78.0	79.5	78.5	0.4
	PIP	Pos 1	0.0	0.5	0.1	0.1
Ring	MCP	Pos 2	98.0	99.2	98.4	0.2
		Pos 1	0.0	0.5	0.1	0.1
	PIP	Pos 2	89.0	90.0	89.4	0.2
Little	MCP	Pos 1	0.0	1.2	0.1	0.2
		Pos 2	98.0	98.5	98.2	0.1
	PIP	Pos 1	0.0	0.5	0.1	0.1
	MCP	Pos 2	89.0	90.6	89.2	0.3
		Pos 1	0.0	0.5	0.1	0.1
	PIP	Pos 2	80.0	81.9	80.4	0.6
	MCP	Pos 1	0.0	0.5	0.1	0.1
		Pos 2	86.0	87.2	86.2	0.3

Table 14: Target and Average Error of the Repeatability Test of Robot Hand (Degree)

Finger Names	Name of Joint	Target (Deg)	Avg (Deg)	Avg Error (Deg)
Thumb	CMC	Pos 1	0.0	0.1
		Pos 2	98.0	99.2
	MCP	Pos 1	0.0	0.1
		Pos 2	76.0	76.1
	DIP	Pos 1	0.0	0.1
		Pos 2	39.0	39.1
Index	MCP	Pos 1	0.0	0.1
		Pos 2	70.0	70.3
	PIP	Pos 1	0.0	0.1
		Pos 2	78.0	78.5
Middle	MCP	Pos 1	0.0	0.1
		Pos 2	98.0	98.4
	PIP	Pos 1	0.0	0.1
		Pos 2	89.0	89.4
Ring	MCP	Pos 1	0.0	0.1
		Pos 2	98.0	98.2
	PIP	Pos 1	0.0	0.1
		Pos 2	89.0	89.2
Little	MCP	Pos 1	0.0	0.1
		Pos 2	80.0	80.4
	PIP	Pos 1	0.0	0.1
		Pos 2	86.0	86.2

#### 4. Conclusions

From the experiment, this anthropomorphic robot hand can grasp selected 32 different objects commonly found in daily life with 10 basic gripping postures and can perform 9 basic gestures. The other gestures that this hand cannot perform use the abduction-adduction movement such as fingers crossed, fig sign, and Vulcan salute. This robot hand can increase grasping posture and hand gesture by adding the abduction-adduction motion with the smallest actuators into the MCP joint of each finger. In addition, the robot hand can grasp an object up to 450 grams. When grasping huge objects, we usually use the cylindrical grasp (power grasp

posture) as in the grasping experiment section. From the above results, it can be found that the motor can sufficiently transmit force and torque to the fingers and fingertips in order to grasp the 450 grams object. By using a Lateral grasp, the anthropomorphic hand can pick up small objects such as keys and utility knives with fingertips. In addition, the proposed robot hand has sufficient force and rigidity to grasp various objects while the cost is lower than other designs. The equations used in the design proven that the structure can move according to the equation with an error value of about 1.6 degrees. In the repeatability experiment, this robot's hand has a maximum error of repeatability of about 1.2 degrees.

We have designed and prototyped an open-source anthropomorphic robotic hand for teleoperated robots with a detailed design process for further developers. We use 3D printing and common components for assembling. The four-bar linkage mechanism is used to mimic the relative motion between DIP and PIP joints same as the human finger, while also reducing the number of motors. We experimentally that our proposed robotic hand design has good repeatability in finger motions and grasping daily objects. This paper explains how to design a robot hand, it can be adjusted to any desired size by using the equation given above.

Design of an Open-Source Anthropomorphic Robotic Hand for Telepresence Robot is available for study and development, which can be found at the following site. <https://github.com/Jittaboontri/Anthropomorphic-Robotic-Hand>

#### Conflict of Interest

The authors declare no conflict of interest.

#### Acknowledgment

We want to give acknowledgements to AI for all projects for the financial support in our research, Institute of Field Robotics (FIBO), and King Mongkut's University of Technology Thonburi (KMUTT) and Fundamental Fund (Basic Research Fund) for supporting fund.

#### References

- [1] J. Trichada, T. Wimornut, N. Tirasuntarakul, T. Choopojcharoen, B. Sakulkueakulsuk, "Design of an open source anthropomorphic robotic finger for telepresence robot," ACM International Conference Proceeding Series, 62-66, 2021, doi:10.1145/3467691.3467704.
- [2] S.C. Jee, M.H. Yun, "An anthropometric survey of korean hand and hand shape types," International Journal of Industrial Ergonomics, 53, 10-18, 2016, doi:10.1016/j.ergon.2015.10.004.
- [3] V. Doroshenko, O. Mul, O. Kravchenko, "Mathematical relations for harmonization with technical and decorative casting nature," Boundary Field Problems and Computer Simulation, 55(December), 44-49, 2016, doi:10.7250/bfps.2016.007.
- [4] P.G. Narasimha-shenoi, Golden ratio in human anatomy, Thesis, Government College Chittur, 2014, doi:10.13140/2.1.2265.9526.
- [5] D. Persaud, J.P. O, "Fibonacci series, golden proportions, and the human biology," Austin Journal of Surgery, 2(5), 1-6, 2015, ISSN : 2381-9030.
- [6] S.A. Powell, A review of anthropomorphic robotic hand technology and data glove based control, Masters Thesis, Virginia Polytechnic Institute and State University 2016.
- [7] M. Controzzi, C. Cipriani, B. Jehenne, M. Donati, M.C. Carrozza, "Bio-inspired mechanical design of a tendon-driven dexterous prosthetic hand," 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'10, (i), 499-502, 2010, doi:10.1109/IEMBS.2010.5627148.
- [8] D. Choi, D.-W. Lee, W. Shon, H.-G. Lee, "Design of 5 D.O.F robot hand

- with an artificial skin for an android robot,” *The Future of Humanoid Robots - Research and Applications*, (January), 2012, doi:10.5772/26282.
- [9] W.S. You, Y.H. Lee, H.S. Oh, G. Kang, H.R. Choi, “Design of a 3D-printable, robust anthropomorphic robot hand including intermetacarpal joints,” *Intelligent Service Robotics*, 12(1), 1–16, 2019, doi:10.1007/s11370-018-0267-8.
- [10] Material property of RA-22AB, [https://www.resinrungle.com/ra\\_22ab.html](https://www.resinrungle.com/ra_22ab.html), Jan. 2022.
- [11] The coefficient of the PTFE tube, <https://www.fluorotherm.com/technical-information/materials-overview/ptfe-properties/>, Jan. 2022.
- [12] Z. Su, K. Inaba, A. Karmakar, A. Das, “Characterization of mechanical property of pla-abs functionally graded material fabricated by fused deposition modeling,” *Proceedings of ASME 2021 Gas Turbine India Conference, GTINDIA 2021*, (December), 2021, doi:10.1115/GTINDIA2021-76025.
- [13] Material price of polymaker, <https://us.polymaker.com>, Jan. 2022.
- [14] Material price of eSUN, <https://www.esun3d.com>, Jan. 2022.
- [15] Specification of Dynamixel servo motor, <https://www.robotis.us/dynamixel-xl430-w250-t/>, Jan. 2022.
- [16] Specification of tendon, <https://www.lazada.co.th/products/1-2-proberos-x4-100m-bluegreenyellowredgrey-pe-4-5-100-thailand-fishing-mall-fishing-line-i2161246830-s7196994582.html>, Jan 2022.

## Analysis and Trend Estimation of Rainfall and Seasonality Index for Marathwada Region

Himanshu Bana\*, Rahul Dev Garg

Geomatics Engineering, Indian Institute of Technology Roorkee-247667, India

### ARTICLE INFO

Article history:

Received: 15 September, 2022

Accepted: 23 December, 2022

Online: 24 January, 2023

Keywords:

Seasonality Index

trend estimation

Marathwada

Temporal Analysis

### ABSTRACT

Droughts are undesirable and highly unwanted form of disasters. It is essential to analyse the cause of such extreme events and act accordingly to pave the way for a sustainable future. The present research work conducts a seasonality and trend analysis of rainfall over the eight districts of Marathwada region. The study is carried out for the last 39 years ranging from 1980 to 2018. The rainfall data pertaining to pre-monsoon season, monsoon season (Kharif), and annual average have been analysed. The trend has been estimated using Sen's slope estimation process along with Mann-Kendal test. It was determined that the all the eight districts of the region show a negative trend in the annual rainfall received. Nanded district showed the largest negative trend in the annual rainfall. Out of eight districts seven districts of the region show a decline in rainfall during the monsoon season. The district of Nanded showed largest decline in the rainfall received during monsoon season. The research work presents the discussion on possible causes of such trends estimated. The research creates a robust foundation of advanced computation techniques for prediction of droughts.

### 1. Introduction

Melting of glaciers, frequent droughts, and increase in regional temperature are some of the climatic changes that are expected to affect the agricultural scenario of the world [1]. Due to these adverse climate changes, it has been predicted by the intergovernmental panel on climate change (IPCC) that these unfavourable events may lead to scarcity of drinking water and water resources [2]. As per the panel this scarcity of water resource might led to drop in per capita freshwater availability. The effect of this drop would be visible by 2025.

Many researchers in the past have indicated that the change in climatic conditions will bring both scarcities of precipitation and increased intensity of precipitation [3-5]. The increased intensity of precipitation will result in intense flooding, flash flooding, and higher run-off during the monsoon. As the run-off will increase the lesser precipitation will percolate. The ground which is bound to negatively affect the ground water recharge will subsequently result in falling of water table and lower volume of water available for anthropometric activities [6]. Prediction of scarcity of water along with prediction of increased intensity of precipitation indicates that climate change will affect the precipitation at both local and regional scales.

The research presented in [7] indicated in their research work that in Asia-Pacific region the agricultural activities are highly dependent on ground water and monsoon. Thus, depleting

water table and less precipitation will adversely affect the cropping system in the region. As the cropping system will be affected it will lay an effect on yield productivity and the net area sown under the principal crops in the region [8].

IPCC has also predicted the probability that the global surface temperature might increase by 5.8°C by the end of year 2021 [2]. Many researchers in the past have worked upon the sensitivity of the crops to the surface temperature [9,10]. The study in [11] focused on accessing the probable impact of temperature rise on the production of wheat in India. They determined through mathematical modelling that a 1oC rise in temperature is sufficient to drastically reduce the production of wheat.

Study of rainfall variation in India is of special interest to researchers for a long time [12]. The research in [13] focused on studying the variations of climatic parameters in different regions of India even before the subject of climate change was prominent. The special interest in studying the rainfall variations comes from the fact that Indian agriculture is entirely dependent on rainfall. If the states of Punjab and Haryana is not considered no state in India has a proper network of canals and channels that can supply water for irrigation to the farmers. Due to unavailability of irrigation infrastructure the farmers in India are dependent on monsoon. If the monsoon performs poorly in any year, the production of Kharif crops gets affected drastically. It is due to these monsoon dependent characteristics of Indian agriculture; it is called 'Gamble on Rains'.

\*Corresponding Author: Himanshu Bana, himanshu\_roorkee1@yahoo.com

The research in [14] focused on determining the rainfall trend in the North Eastern states of India. They focused on determining the trend present in the North Eastern states because these states suffer due to scarce as well as heavy rainfall [15]. The study in [16] concluded that due to inadequacy of the irrigation system a decreased rainfall results in poor agricultural production while an increased rainfall always poses a certain danger of flooding due to Brahmaputra breaking its banks. Researchers in the past also indicated that monsoon presents a decreasing trend in the states of Chhattisgarh, Jharkhand, and Kerala [17]. In recent years, the monsoon in India is weakened by the El Nino Southern Oscillations (ENSO) especially in the year 2009, 2015, and 2017 [18]. The ENSO negatively affects the monsoon over India. This negative effect causes less than normal rainfall during phases of El Nino.

The Marathwada region of Maharashtra is a drought prone area [19]. Latur and Osmanabad districts of the region are some of the worst affected regions of the country [14,20]. In 2016, numerous full capacity trains only with water wagons were ferried to Latur to meet the water scarcity of the district [10]. The year of 2016 was not the first time a Latur district from Marathwada region, has suffered from severe water shortages in. Latur has faced droughts in 1980s as well in 1990 [21]. However, the event of scarcity of water in April 2016 was an extreme event. The time of the study is selected from 1980 because the region has shown more susceptibility for getting affected by the drought since then. The research work thus, tries to determine and evaluate the trend present in the rainfall received by the districts of Marathwada region through a time-series analysis for the years 1980-2018.

## 2. Research Method

The region selected for the study is Marathwada as shown in Figure 1. The Marathwada region is a group of districts located in south western region of state of Maharashtra. The region is comprised of districts namely Beed, Latur, Parbhani, Hingoli, Jalna, Aurangabad, Osmananbad, and Nanded. The region lies near to the northern ranges of western Ghant. The location of the study region has been shown in the map. The region was earlier known for its sugarcane and cotton production. However, the region has started to witness increased frequency of below

normal rainfall during monsoon which has reduced the sugarcane cultivation in the region. Further, the area becomes area of interest for the study because the district of Latur recently faced one of the worst water crises in history of Independent India [20].

## 3. Data

The present study is based on data recorded at 8 stations in the Marathwada region. The period of the data is from 1980 to 2018 (Last 39 years). The data was procured from the India Meteorological Department (IMD). The rainfall data used in this research work was recorded in the stations in the form of direct observation and was subjected to standard normal homogeneity test for homogenization. The data contained no missing values. The trend in the rainfall in the selected districts was estimated on an annual, pre-monsoon, and Kharif season (main sowing season in India that begins in June. The season begins with advent of South West Monsoon making a landfall at Kerala). The pre monsoon months were selected as March, April, and May. The Kharif season was selected as June-July-August, and September.

## 4. Analysis

For the study seasonality index (SI), standard deviation (SD), coefficient of variance (CV), Sen Slope and Mann-Kendall test were utilized. The seasonality index helps in determination of contrast in the rainfall regime. This process is done by utilization of rainfall monthly distribution data. In other words, the seasonality index helps in identification of monthly rainfall variability [22]. The seasonality is a function of mean of monthly rainfall and mean annual rainfall. The seasonality index is computed as:

$$SI = \frac{1}{A} \sum_n^{12} \left| X_n - \frac{A}{12} \right| \quad (1)$$

where,  $X_n$  is the rainfall calculated for the  $n$ th month.  $A$  is the total annual rainfall. Theoretical variations in the seasonality index can be from 0 to 1.83. If all the months in a year records equal rainfall, than the SI becomes zero. If all the rainfall occurs in one month than the SI becomes 1.83 [23]. The SI also suggests changes in rainfall pattern [24]. The rainfall regimes associated with the different values of the seasonality index has been shown in the table 1.

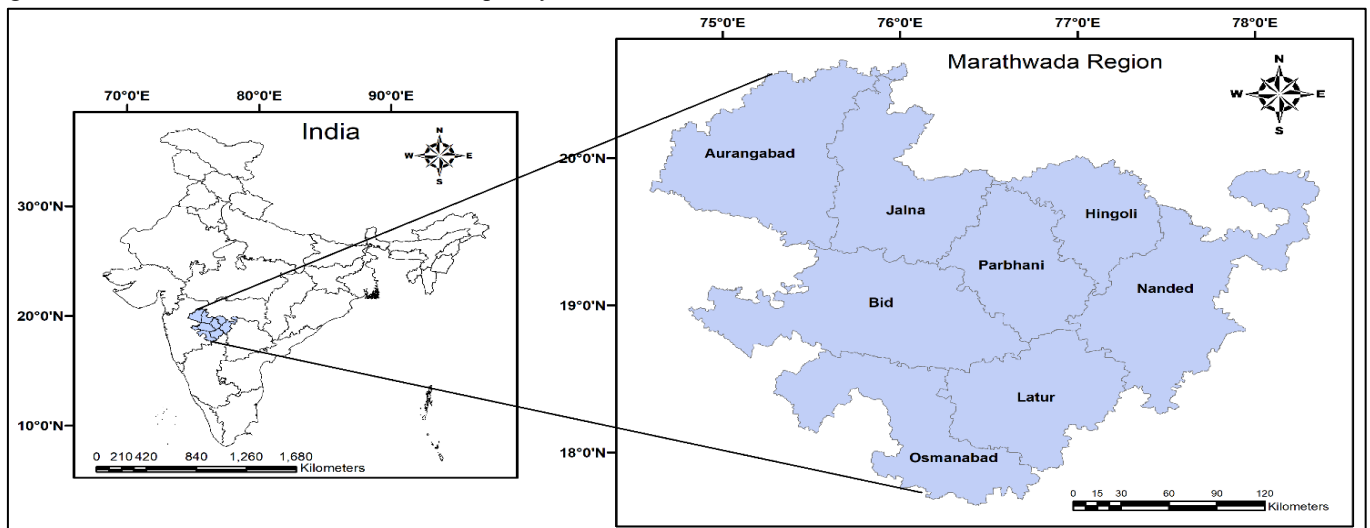


Figure 1: Study Area: Marathwada region

Table 1: Rainfall regimes and associated SI (Walsh & Lawer, 1981)

Regimes	Seasonality Index (SI)
Very Equable	Less than or equals 0.19
Equable, but with a definite wetter season	0.20-0.39
Rather Seasonal with short drier season	0.40-0.59
Seasonal	0.60-0.79
Marked Seasonal with long drier season	0.80-0.99
Mostly rain in 3 months or less	1.00-1.19
Extreme, almost rain in 1 to 3 months	Greater than or equals 1.20

Along with the identification of rainfall regime, SI indicates towards soil and vegetation characteristics along with hydrological stress in the region. Rainfall trend analysis can be performed by using different available parametric and non-parametric methods [25]. These analyses are in general meant to analyse the trend present in long term dataset. However, these techniques are also used to determine trend in short-term data series [26]. These short-term data series can be truncated to ten data points. The restriction associated with the use of parametric test in the trend determination is that the data points in the time series should follow a distribution. Non-parametric tests do not pose such restrictions and are minimally affected by any outliers present in the dataset. In the present work the trend present in the rainfall data was analysed using Mann-Kendal test and the Sen-slope estimates. The slope estimates were determined using Sen-slope estimation process. The Mann-Kendall tests are widely utilized for determination of trends which are monotonous in nature in the non-cyclic data sets [27]. The wide popularity in determining the trend present in the rainfall data using Mann-Kendall test is due to because Mann-Kendal test does not require a particular distribution. Another important characteristic of Mann-Kendal test is that it is least sensitive towards in homogeneity present in the time series [28]. The time series is thus assumed to obey the following model.

$$x_i = f(t_i) + \epsilon_t \tag{2}$$

where,  $f(t)$  is the monotonic decreasing or increasing function of time; the residual is represented by  $\epsilon_t$ . It is also assumed further that the variance of the distribution is constant in time. Further, it is also assumed that the autocorrelation in the data set is zero [29].

The estimate of slope which is denoted by  $Q$  is calculated by the following process [30].

Firstly, the slope between all pairs of data values is determined. The procedure for the same is as follows.

$$Q_i = \frac{x_j - x_k}{j - k} \text{ for } i=1,2,3, \dots, k \tag{3}$$

where,  $x_j$  and  $x_k$  are the data values at time  $j$  and  $k$  ( $j > k$ ).

In time series that contains  $n$  values the  $N$  would be determined using

$$N = n(n-1)/2 \tag{4}$$

Where,  $N$  is the no of iterations and  $T$  is total observations.

The  $N$  values are arranged in ascending order before the application of  $T$  in the equation presented below. Estimator of Sen's slope is representative of the median of these  $N$  values of  $Q_i$

The  $Q$  is thus given by,

$$\text{If } N \text{ is odd, } Q = T_{(N+1)/2} \tag{5}$$

$$\text{If } N \text{ is even, } Q = \frac{1}{2} (T_{N/2} + T_{N/2+1})$$

Normal distribution is applied for determination of two-sided confidence interval related to the estimation of the slope. In the dataset of the time series the downward or the decreasing trend is indicated by the negative value of  $Q$ , while an upward or increasing trend is determined by positive value of  $Q$ . The Mann-Kendal test null hypothesis assures that the data series had normal distribution. A significance level of 0.001 has been used for the testing-module. These characteristics correlate to the fact that a probability of 0.01% exists for the randomness in the time series data. Similarly, if the significance level is kept as 0.05 then it is assumed that there exists a 5% of probability that the values of the time series are from random distribution. The Mann-Kendall test statistics  $S$  is determined using the formula [31].

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sgn}(x_j - x_k) \tag{6}$$

$$\text{sgn}(x_j - x_k) = \begin{cases} 1 & \text{if } x_j - x_k > 0 \\ 0 & \text{if } x_j - x_k = 0 \\ -1 & \text{if } x_j - x_k < 0 \end{cases} \tag{7}$$

The number of data points in the time series under consideration is denoted by  $n$  and  $x_k, x_j$  are the rainfall sum in the year  $j$  and  $k$  respectively. Here, the  $j^{\text{th}}$  value is greater than the  $k^{\text{th}}$  value. The present work fixes the significance level as 0.05 for the test. Since the number of data points would be greater than 10 the distribution of  $S$  was approximated using a normal distribution. The estimation of trend line was done using linear regression method [32]. In this research work slopes that are of significant nature at significance level of 5% has been marked with\*. Slopes that are not of significant nature are not marked with any sign in the result tables.

## 5. Results

The results of the present study are as follows:

### 5.1. Seasonality Index

The results of the seasonality index (SI) have been presented in the table 2. The SI reveals that Beed, Latur and Osmanabad face frequent long drier season (SI ranging between 0.8 - 0.99, are marked by red colour cells). In these districts the long drier season occurred for 11, 13 and 9 times respectively in 39 years. Parbhani, Hingoli, Jalna, Aurangabad, and Nanded also face long drier season but in these districts the occurrence of long drier season is less as compared to the rest of three districts. The district of Nanded faced least number of long drier seasons in last 39 years. The SI in the Nanded district for most of the years is greater than 1 which indicates that the district receives rains in



almost three months or less. The district of Latur faced two consecutive long drier seasons in last 39 years. The first occurred from 1985 to 1987 and the second occurred from 2013 to 2015.

Further, the district of Latur faced long drier season in alternate years from 1987 to 1992 and from 2002 to 2006.

Table 2: Seasonality Index for the last 39 years for Districts of Marathwada Region

Year	Beed	Parbhani	Hingoli	Latur	Jalna	Aurangabad	Osmanabad	Nanded
1980	1.21	1.07	1.3	1.19	1.25	1.24	1.25	1.24
1981	0.85	1.15	1.03	1.02	1.24	0.9	0.94	1.16
1982	1.05	1.06	0.94	1.04	0.9	0.92	0.95	1.05
1983	1.17	1.04	1.19	1.15	0.92	1.24	1.12	1.17
1984	1.11	1.04	1.11	1.07	1.24	1.1	1.11	1.09
1985	1.09	1.09	1.07	0.97	1.1	1.15	1.05	1.03
1986	1.1	1.28	1.14	0.87	1.15	1.22	1.11	1.07
1987	0.99	1.18	1.12	0.92	1.22	1.05	0.94	1.03
1988	1.26	0.98	1.24	1.27	1.05	1.09	1.19	1.16
1989	1.14	1.39	1.2	1.11	1.09	1.2	1.08	1.22
1990	0.98	1.06	1.03	0.97	1.2	0.95	0.92	0.97
1991	1.27	0.95	1.38	1.12	0.95	1.33	1.09	1.25
1992	1.14	1.03	1.11	0.97	1.33	1.15	1.01	1.13
1993	0.95	0.86	1	1.04	1.15	1.01	1	1.05
1994	0.88	1.24	1.16	1.05	1.01	0.92	1.02	1.11
1995	0.93	0.78	0.94	0.92	0.92	1.1	0.9	1.02
1996	1.17	1.08	1.2	1.14	1.1	1.12	1.1	1.16
1997	0.75	1.13	0.91	0.71	1.12	0.87	0.76	0.78
1998	1.1	1.31	1.04	1.06	0.87	1.07	1.12	1.08
1999	1.05	1.2	1.14	0.95	1.07	1.09	1	1.13
2000	1.2	1.17	1.37	1.13	1.09	1.12	1.11	1.2
2001	1.2	1.23	1.23	1.14	1.12	1.13	1.17	1.2
2002	1.14	1	1.26	0.94	1.13	1.17	1	1.25
2003	1.06	1.13	1.25	1.18	1.17	1.04	1.01	1.18
2004	0.89	1.13	1.07	0.89	1.04	1.12	0.83	1.01
2005	1.16	1.31	1.12	1.09	1.12	1.2	1.16	1.11
2006	0.95	1.21	1.17	0.96	1.2	1.11	1	1.19
2007	1.28	0.91	1.31	1.26	1.11	1.22	1.15	1.23
2008	1.05	1.14	1.24	1.13	1.22	1.24	1.14	1.22
2009	0.84	1.23	0.89	1.02	1.24	1.05	0.92	1.05
2010	1.05	1.17	1.12	1	1.05	0.94	1.01	1.07
2011	1.18	1.02	1.28	1.3	0.94	1.25	1.15	1.26
2012	1.15	0.98	1.17	1.16	1.25	1.17	1.13	1.15
2013	1.01	0.99	1.07	0.96	1.17	1.09	1.01	1.01
2014	0.9	1.13	1.1	0.85	1.09	0.91	0.79	1.08
2015	0.87	1.16	1.01	0.81	0.91	1.05	0.82	0.87
2016	1.12	1.27	1.12	1.07	1.05	1.2	1.09	1.09
2017	1.14	1.2	1.15	1.1	1.2	1.15	1.1	1.1
2018	1.18	1.11	1.33	1.05	1.15	1.23	0.95	1.19

### 5.2 Rainfall Trend

The trend in the annual rainfall for the various district of Marathwada regions has been shown in the Table 3.

Table 3: Annual Rainfall Trend Analysis for the last 39 Years (1980-2018)

District	Mean (mm)	Standard Deviation (mm)	Sen's Slope (mm/year)	Mann Kendall (mm/year)
Beed	745.33	187.45	-1.962	-0.75
Parbhani	872.20	246.12	-6.090	-1.62
Hingoli	903.80	250.53	-4.771	-1.28
Latur	814.83	210.11	-0.460	-0.12
Jalna	747.47	167.87	-1.703	-0.53

Aurangabad	680.276	156.01	-1.787	-0.70
Osmanabad	737.724	193.05	0.021	0.00
Nanded	985.91	311.88	-5.60*	-1.65

\*Statistically significant at 5%, Significance tested using MK Test.

The mean rainfall in the Beed district was observed as 745.33 mm from 1980 to 2018. The annual rainfall in Beed district years shows a negative trend for the last 39 in the annual rainfall. The Sen's slope determined for the annual rainfall received by the district in the last 39 years is -1.962 mm/Year. The mean rainfall in the Parbhani district was observed as 872.20 mm from 1980 to 2018. The annual rainfall of Parbhani for the

last 39 years shows a negative trend. The Sen's slope determined for the annual rainfall received by the district in the last 39 years is -6.090 mm/Year. The mean rainfall in the country's one of most severely and frequently drought affected district Latur was observed as 814.83 mm from 1980 to 2018. The annual rainfall trend in Latur district shows a negative trend for the last 39 years. The Sen's slope determined for the annual rainfall received by the district in the last 39 years is -4.771 mm/Year. The mean rainfall in the Nanded district was observed as 985.91 mm from 1980 to 2018. Out of the 8 selected districts, Nanded received highest mean annual rainfall during the last 39 years. The annual rainfall in Nanded district for the last 39 years shows a negative trend. The Sen's slope determined for the annual rainfall received by the district in the last 39 years is -5.60 mm/Year which was statistically significant negative trend.

The pre-monsoon season holds special importance in the Marathwada region known for its water intensive crops such as sugarcane [33]. Industries in and around the Marathwada region are also known to use water intensively. These industries include the sugar mills and the cotton dying industries. Real-estate activities are also on the rise in the region [34]. Traditional construction approach adopted in the region requires excessive use of water for curing of concrete and the wall plaster. The demand for the sugarcane increases in summer season due to fresh juice stalls and sugar mills boost their production to stock the sugar for the upcoming festive seasons [35].

Farmers use excessive water in the fields during pre-monsoon season in order to increase the brix content of the crop and inter nodal gap in sugarcane [4]. Farmers in the region typically use drenching method of irrigation which also accounts for loss of precious water reserves [36]. Such anthropogenic activities require intense water and may pose a certain threat to the water availability in this already drought susceptible region of the Maharashtra state. From table 4, it is evident that Pre-monsoon rainfall trend is positive for all the 8 districts under consideration. The districts of Jalna and Aurangabad show a positive trend in the pre-monsoon rainfall, but the trend determined is minuscule. Largest positive trend in the pre-monsoon rainfall was observed for the Osmanabad district. The increasing positive trend is beneficial for the water intensive crops sown in the district.

Table 4: Pre-monsoon Rainfall Trend Analysis of last 39 Years (1980-2018)

District	Mean (mm)	Standard Deviation (mm)	Sen's Slope (mm/yr)	Mann Kendall (mm/yr)
Beed	31.86	29.81	0.295	0.69
Parbhani	27.67	28.30	0.192	0.77
Hingoli	22.02	24.33	0.108	0.31
Latur	43.82	33.28	0.393	0.95
Jalna	20.76	24.23	0.017	0.08
Aurangabad	16.55	23.58	0.049	0.55
Osmanabad	36.92	29.72	0.485	1.63
Nanded	29.82	31.34	0.254	1.27

However, the positive trend determined for the pre-monsoon rainfall in the districts was not statistically significant.

The monsoon season is the time to sow the Kharif crops. Kharif crops are known to be water intensive. Farmers in the

region have a strong affinity for the sowing water intensive crops in the region. Scarcity of water in the germination and early development stage results in osmotic stress [37]. Such stress exploits the growth of the crop. Farmers are known to sow crops like cotton and groundnut in the early monsoon season. Water intensive crops like sugarcane are sown in the middle of the monsoon season so that the crop can be harvested by March to May [38]. Therefore, good monsoon is essential for the Marathwada region from agricultural perspective. Table 5 depicts that the highest mean rainfall in the monsoon season was received by Nanded district. The district of Aurangabad and Osmanabad received mean rainfall of 557.71 mm and 571.12 mm, respectively. Latur which is one of the drought susceptible districts of the country received a mean rainfall of 656.82 mm in the monsoon season. Sen's slope estimate shows the negative trend in the rainfall received by the districts during monsoon.

The district of Beed shows a negative trend of -0.956 mm/Year. The district of Parbhani shows a negative trend of -2.809 mm/Year. The district of Nanded shows the highest negative trend of -3.996 mm/Year followed by the district of Hingoli which shows a negative trend of -3.154 mm/Year. Only the district of Latur shows a positive trend of 0.458 mm/Year for the rainfall received in monsoon season. The negative trend in seven out of eight districts of region is bad from the perspective of sugarcane producers of the region. Figure 2 shows the trend obtained for the rainfall received by the districts in the monsoon season.

Table 5: Kharif Rainfall Trend Analysis of last 39 Years (1980-2018)

District	Mean (mm)	Standard Deviation (mm)	Sen's Slope (mm/yr)	Mann Kendall (mm/yr)
Beed	609.66	179.318	-0.956	-0.34
Parbhani	744.89	229.24	-2.809	-0.77
Hingoli	792.54	227.45	-3.154	-0.90
Latur	656.82	191.40	0.458	0.19
Jalna	620.59	152.30	-1.385	-0.60
Aurangabad	557.71	131.31	-1.669	-0.68
Osmanabad	571.12	167.68	-0.266	-0.05
Nanded	836.07	274.27	-3.996	-1.23

From the analysis of annual and monsoon rainfall received by the districts it is evident that the scarcity of the rainfall in the region is on the rise. Seven districts of region showed a negative trend in the annual rainfall received while seven out of eight districts of the region showed a negative trend in the rainfall received in the monsoon season.

## 6. Discussion

The Thar Desert and adjoining areas of central and northern subcontinents heat up during the summers. This creates a void. To fill up the void the air from the Indian ocean rush into the mainland. The air is laden with moisture picked up from the ocean surface. The Himalayas regulates the air-flow and prevent its influx into the central Asia. As the wind rises, precipitation occurs and India receives rainfall [39]. This rainfall season is also known as the Southwest (SW) monsoon. Marathwada region of Maharashtra state receives the SW monsoon. The period of SW monsoon starts from June and

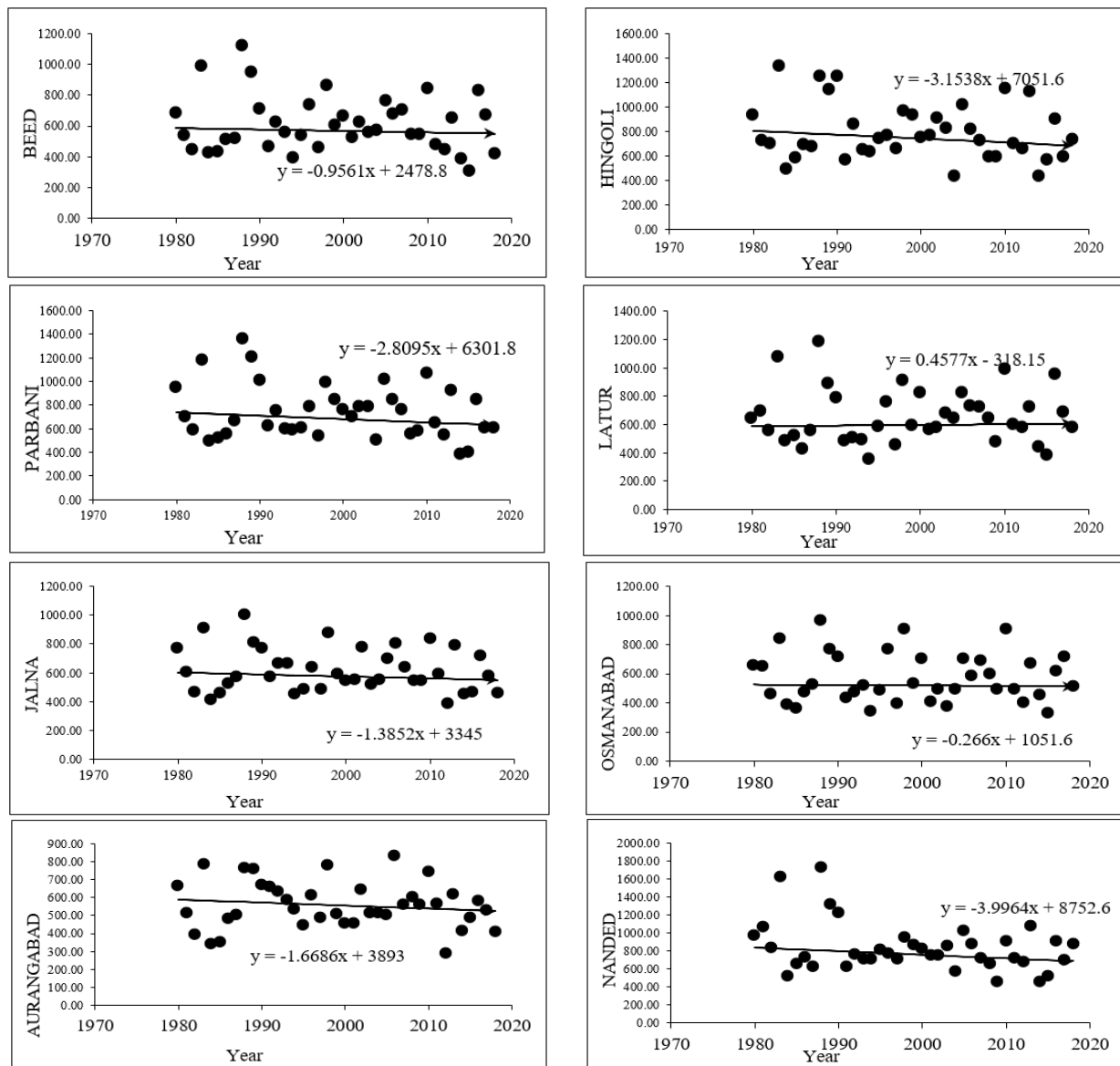


Figure 2: Monsoon rainfall trend from 1980 to 2018 for districts of Marathwada

ends in early to mid of October. The SW monsoon is considered as the principal rainy season in India. Nearly, the whole country receives rainfall during this period. The Southwest monsoon accounts for nearly 75% of rainfall in the country and thus agrarian activities are dependent on it [40].

The SW monsoon is important from India's agricultural perspective. India does not possess any significant irrigation network. Only the states of Punjab and Haryana have proper irrigation infrastructure in place [41,42] Agriculture of rest of the country either survives on monsoon or depends upon the ground water source [43]. It is due to this high dependency on the monsoon, Indian agriculture is often referred as gamble on rains [44].

The characteristics of Indian agriculture is such that the farmers' despite being heavily dependent on rains sow water intensive crops on large scale in the Kharif season (monsoon). Such crops are Paddy and sugarcane. Paddy is prominently grown in the central and eastern regions of India such as Chhattisgarh and West Bengal while Sugarcane is a prominent

Kharif crop of Maharashtra which belongs to western region of India. These crops are sensitive to climate change as it causes rise in temperature, and water scarcity [6]. These effects along with ENSO bring in uncertainty over the amount of rainfall [45]. Thus, climate change poses a certain threat to the Indian agriculture.

From the analysis, it is evident that amount of rainfall in the districts of Marathwada is decreasing. Declining trend was determined for the annual rainfall in all the districts. Farmers in the region are known to produce sugarcane. The sugarcane demands for extensive irrigation for increased brix content, grass weight, and larger node to node distance. With the decreasing rainfall in the region, the farmers are facing a loss in the production. The average productivity of sugarcane crop in the Marathwada region is 50 tonnes per acre while the average productivity of the crop in the state of Maharashtra is 80 tonnes per acre [46]. Thus, it can be concluded that decline in the rainfall in monsoon season (Kharif) is projecting its effect on the productivity of the crop in the region. Absence of irrigation infrastructure in the region results in utilization of ground water by the farmers for the irrigation purposes. This activity further

adds up to the woes of the farmers itself. With decline in rainfall the percolation of water during the rainy season also declines which restricts the recharge of the water table [47].

Utilization of ground water in such cases only degrades the water table. Farmers who are not sowing sugarcane are also facing the effects of decreased rainfall. In recent years farmers of the Latur district had to re-sow their crops because of the long drier season [48].

It is being observed that along with annual rainfall the monsoon rainfall is also depicting a negative trend. The trends are huge for districts like Hingoli, Parbhani, and Nanded. Latur and Osmanabad are the districts that are already receiving less amount of rainfall. In such cases, when the amount of rainfall received is declining the farmers should shift from the sugarcane crop to less water intensive crops. Agriculture activity of such crops is an anthropogenic activity that is adding to water crisis of the region. The seasonality index indicates that Latur frequently faces long drier seasons.

The inferior quality of soil in the Udgir, Ausa and Ahmedpur taluka of the Latur district becomes hard during the long drier seasons which along with steep terrain of the region restrict the percolation of water during the rainy season thus further restricting the ground water recharge. Agriculture is not the only anthropogenic activity that are creating water crisis in the region. Jhum style or the shifting style of agriculture is also prominent in the region. Illegal encroachment of forest land is common in the region. The land is cleared by burning the vegetation present [1]. The burning of vegetation leads compressed temperature difference between the land and the Indian Ocean [49]. This reduction in the temperature between land and sea restricts the draft of air from the ocean that further decreases the rainfall amount. Sugarcane crop not sold to the sugar mills is crushed down in makeshift factories to produce Jaggery and country liquor.

The bagasse is used as a bio-fuel for production of heat needed for making jaggery and liquor. The burning of such bio-fuels is a prominent activity in the region [50]. Although bagasse is low in sulphur content but burning it on a large scale releases ample amount of sulphur dioxide, greenhouse gases and nitrogen oxides into the atmosphere [51]. These emissions further reduce the difference between land and ocean temperature and thus acts as a weakening force for the monsoon system. The warming of Indian Ocean is also leading to rainfall woes in India. The warming of Indian Ocean is leading to decrease the difference between ocean and land temperature. This further reduces the rainfall received by the region. Warming of Indian ocean also results in the occurrence of extreme events [52].

The extreme rainfall event in the districts of Latur and Nanded has been credited to the warming of Indian Ocean [53]. Occurrence of such extreme rainfall events in the backdrop of reducing monsoon might lead to sequence of catastrophic events such as loss of livestock, poverty, and agrarian crisis.

## 7. Conclusion and Future Scope

Drought is an extreme event which has exponentially grown in numbers affecting countless lives and resources' availability. The prominent challenge is experienced by Suryaputra countries

falling under International Solar Alliance like India, Brazil, Australia and South Africa. This creates an alarming issue for a necessity of studies that can assist in eradicating such treacherous events. The region of Marathwada has been a pivotal region prone to such events. All the districts of the Marathwada region of the Maharashtra state are witness a decline in the annual rainfall. The decline in the annual rainfall was largest for the Nanded and Parbhani district. The decline of the annual rainfall in the Nanded district was found to be statistically significant. Out of eight districts seven districts of the region have witnessed a decline in the monsoon rainfall over the last 39 years.

The seasonality index calculated indicates that Latur, Beed, and Osmanabad are the drier districts of the region which receives rainfall in more than 3 months. The negative rainfall trend observed might pose a threat to the highly monsoon dependent agriculture of the region. The farmers of the region therefore should migrate from sowing water intensive crops to less water intensive crops such as Sorghum and pearl millet. Sugarcane can also be replaced with less water intensive mandarin which is a less water intensive cash crop. In cases where the farmers are not able to shift to other crops irrigation management system such as irrigation through drip irrigation and rain pipes should be implemented. If negative trend in the rainfall is not effectively managed through change and control of anthropogenic activities the region of Marathwada might enter advanced phases of agrarian crisis which might also lead to the collapse of the agriculture system of the districts comprising it.

The study presented rainfall and seasonality trends which portrayed the topographical and climatic conditions of the districts of Marathwada region. Based on the analysed dataset of 39 years, it can very well be inferred that a spatial time-series analysis yields fruitful information regarding the aspects, characteristics and trend analysis which acts as dominant prerequisites for advanced computation techniques to be deployed for analysis and prediction of droughts in the upcoming years.

## References

- [1] K. Gabhiye, C. Mandal, *Agro-Ecological Zones, their Soil Resource and Cropping Systems*. Nagpur: National Bureau of Soil Survey and Land Use Planning, 2000.
- [2] IPCC, *The physical science basis. The contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. New York: Cambridge University Press, 2007.
- [3] M. Dore, "Climate Change and Changes in global precipitation" *Environmental International*, **31**, 1167-1181, 2005.
- [4] S. Manwar, P. Vadiya, "Characterization and classification of sugarcane growin soils of Latur district", *Annals of Plant and Soil Research*, **17**(5), 373-377, 2015.
- [5] A, Saini, N. Sahu, P. Kumar, S. Nayak, W. Duan, R. Avtar, S. Behera, "Advanced Rainfall Trend Analysis of 117 Years over West Coast Plain and Hill Agro-Climatic Region of India", *Atmosphere*, **11**(11), 1225, 2020, 10.3390/atmos11111225.
- [6] D.Y. Gumel, A.M. Abdullah, A.M. Sood, R.E. Elhadi, M.A. Jamalani, K.A.A.B. Youssef, "Assessing Paddy Rice Yield Sensitivity to Temperature and Rainfall Variability in Peninsular Malaysia Using DSSAT Model". *International Journal of Applied Environmental Sciences*, **12**(8), 1521-1545, 2017.
- [7] M. Parry, O. Canziani, J. Palutikof, P.V.D. Linden, C. Hanson, *Aia Climate Change 2007: Impacts, adaptation and vulnerability*. In : Fourth Assessment report of the intergovernmental panel on climate change. Cambridge: Cambridge University Press, 2007.
- [8] A. Saini, N. Sahu, W. Duan, M. Kumar, R. Avtar, M. Mishra, S. Behera, "Unraveling Intricacies of Monsoon Attributes in Homogenous Monsoon

- Regions of India", *Frontiers in Earth Science*, 10, 1–17, 2022, 10.3389/feart.2022.794634.
- [9] M.A. Semenov, "Impacts of Climate Change on Wheat in England and Wales", *Journal of the Royal Society Interface*, 6, 2008, 343-350. doi:10.1098/rsif.2008.0285
- [10] S. Osmani, P. Patil, "Drought response and relief by Jaldoot Express: A case study of Latur drought", *Zenith IJMR*, 9(6), 224-236, 2019.
- [11] A. Kumar, A. Singh, "Climate Change and its Impact on Wheat Production and Mitigation through Agroforestry Technologies", *International Journal of Environmental Sciences*, 5(1), 73-90, 2014.
- [12] O. Dhar, B. Parthasarathy, "Trend analysis of annual Indian rainfall", *Hydrological Science*, 26, 257-260, 1975.
- [13] K. Krishnamurthi, Y. Ramanathan, "Sensitivity of the monsoon onset to differential heating", *Journal of atmospheric science*, 39, 1290-1306, 1982.
- [14] D. Duhan, A. Pandey, "Statistical analysis of long term spatial and temporal trends of precipitation during 1901-2002 at Madhya Pradesh", *Atmospheric Research*, 122, 136-149, 2013.
- [15] V. Kumar, S.K. Jain, Y. Singh, "Analysis of long-term rainfall trends in India", *Hydrological Sciences Journal*, 55(4), 484-496, 2010, doi:10.1080/02626667.2010.481373.
- [16] A. Gupta, "Flood and Floodplain management in North East India: An Ecological Perspective", 1st International Conference on Hydrology and Water Resources in Asia Pacific Region, Kyoto: Hydrology and Water Resources, 1-10, 2003.
- [17] S. Swain, M.K. Verma, M. Verma, "Analysis of Change in Annual Rainfall for Raipur district", *IJERT*, 3(20), 1-10, 2015.
- [18] I. Roy, R.G. Tedeschi, M. Collins, "ENSO teleconnections to the Indian summer monsoon under changing climate", *International journal of climatology*, 39(6), 3031-3042, 2019.
- [19] A. Kulkarni, S. Gadgil, S. Patwardhan, "Monsoon variability, the 2015 Marathwada drought and rainfed agriculture". *Current Science*, 111(7), 1182-1193, 2016.
- [20] SANDRP, Latur Drinking Water Crisis highlights absence of Water Allocation Policy and Management, Retrieved from South Asia Network on Dams, Rivers and People, 2016 <https://sandrp.in/2016/04/20/latur-drinking-water-crisis-highlights-absence-of-water-allocation-policy-and-management/>
- [21] D. Kolekar, V. Vanama, Satellite based Drought Assessment Over Latur, India Using Soil Moisture Derived From SMOS. ISPRS TC V Mid-term Symposium, Geospatial Technology – Pixel to People. Dehradun: ISPRS, 2018, doi:10.5194/isprs-archives-XLII-5-421-2018.
- [22] E. Kanellopoulou, "Spatial distribution of rainfall seasonality in Greece", *Weather*, 57, 215-219, 2002.
- [23] S. Ingle, S. Patil, N. Mahale, Y. Mahajan, "Analyzing rainfall seasonality and sNorth Maharashtra", *Environmental Earth Sciences*, 77, 651-662, 2018.
- [24] R. Walsh, D. Lawer, "Rainfall seasonality: description, spatial patterns and change through time", *Weather*, 36, 201-208, 1981.
- [25] R. Yadav, S. Tripathi, G. Pranuthi, S. Dubey, "Trend analysis by Mann-Kendall test for precipitation and temperature for thirteen districts of Uttarakhand", *Journal of agrometeorology*, 16(2), 164-171, 2014.
- [26] I. Ahmad, D. Tang, T. Wang, M. Wang, B. Wagan, "Precipitation Trends over Time Using Mann-Kendall and Spearman's rho Tests in Swat River Basin", *Advances in meteorology*, 2015, 1-15, 2015, doi:10.1155/2015/431860
- [27] H. Tabari, S. Marofi, M. Ahmadi, "Long-term variations of water quality parameters in the Maroon river", *Environmental monitoring assess*, 177, 273-287, 2011.
- [28] N. Karmeshu, Trend Detection in Annual Temperature & Precipitation using the Mann Kendall Test – A Case Study to Assess Climate Change on Select States in the Northeastern United States. Pennsylvania, 2012.
- [29] S. Yue, P. Pilon, B. Phinney, G. Cavadias, "The influence of autocorrelation on the ability to detect trend in hydrological series", *Hydrological processes*, 16(9), 1807-1829, 2002.
- [30] S. Shahid, "Trends in the extreme rainfall events in Bangladesh", *Theoretical Application of Climatology*, 104, 489-499, 2011.
- [31] R. Gilbert, *Statistical methods for environmental pollution monitoring*. New York: Van Nostrand Reinhold, 1987.
- [32] F. Wang, W. Shao, H. Yu, G. Wang, X. He, D. Zhang, G. Kan, "Re-evaluation of the Power of the Mann-Kendall Test for Detecting Monotonic Trends in Hydrometeorological Time Series", *Frontiers in Earth Science*, 8, 1-12, 2020, doi:10.3389/feart.2020.00014
- [33] GWP, Droughts and Sugar Industry in Maharashtra – Are We Learning from History? New Delh: Global Water Partnership, 2016.
- [34] S. Sandbhor, "Analysis of Behaviour of Real Estate Rates in India-A Case Study of Pune City", *International Journal of Economics and Management Engineering*, 7(8), 2465-2570, 2013.
- [35] R. Singh, "Sugarcane marketing systems in India", *Sugar Technology*, 13(4), 1-10, 2011.
- [36] M. Sabesh, M. Ramesh, H. Prakash, G. Bhaskaran, "Is there any shift in cropping pattern in Maharashtra after the introduction of Bt Cotton", *Indian society for cotton improvement*, 6(1), 63-70, 2014.
- [37] C. Pote, A. Kale, "Effect of Osmotic Stress on Sugarcane (*Saccharum officinarum* L.) Growth and Physiology", *International Journal of Current Microbiology and Applied Sciences*, 8(12), 1472-1481, 2019.
- [38] R. Garkar, *Sugarcane Breeding*. Central Sugarcane Research Station, Padegaon, 2017.
- [39] M. Rajeevan, D. Pai, R. Kumar, B. Lal, "New statistical models for long-range forecasting of southwest monsoon rainfall over India", *Climate Dynamics*, 2-17, 2007, doi:10.1007/s00382-006-0197-6
- [40] S. Sasane, "Impact of south west monsoon on crop yield: a statistical analysis", *International Interdisciplinary Seminar on Geographical and Historical Perspective of Global Problems*, 1-10, 2017.
- [41] J. Skutsch, J. Rydzewski, Review of research and development needs in irrigation and drainage, Romw: FAO, 2001.
- [42] R. Jain, P. Kishore, D. Singh., (2019). "Irrigation in India: Status, challenges and options", *Journal of soil and water conservation*, 18(4), 2455-2459, 2019.
- [43] K. Kumar, "Climate impacts on Indian agriculture", *International journal of climatology*, 24(11), 1375-1393, 2004.
- [44] S. Gadgil, "The Indian Monsoon". *Resonance*, 11(8), 8-15, 2006.
- [45] K. Tamaddun, "Effects of ENSO on Temperature Precipitation and Potential Evapotranspiration of North India's Monsoon: An Analysis of Trend and Entropy", *Water*, 11(2), 1-21, 2019, doi:10.3390/w11020189
- [46] P. Upreti, A. Singh, "An Economic Analysis of Sugarcane Cultivation and its Productivity in Major Sugar Producing States of Uttar Pradesh and Maharashtra", *Economic Affairs*, 62(4), 711-718, 2017.
- [47] A. Dias, R. Dhawde, N. Surve, A. Weinberg, T. Birdi, N. Mistry, "Impact of climate changes on water availability and quality in the state of maharashtra in western India", *Asian Jr. of Microbiol. Biotech. Env. Sc*, 17(4), 1071-1081, 2015.
- [48] N. Jamwal, Maharashtra Farmers Fear Loss of Kharif Harvest, Blame Met Department. *The Wire*, 2017.
- [49] B. Singh, O. Singh, "Study of Impacts of Global Warming on Climate Change: Rise in Sea Level and Disaster Frequency", *Global warming Impacts and Future Perspective*, 1-10, 2012, doi:10.5772/50464.
- [50] S. Kulkarni, "Development of efficient furnace for jaggery making", *International Journal of Recent Scientific Research*, 9(5), 26563-226565, 2018.
- [51] J. Halofsky, B. Harvey, "Changing wildfire, changing forests: the effects of climate change on fire regimes and vegetation in the Pacific Northwest, USA", *Hydrobiologia*, 16(4), 1-10, 2020.
- [52] M. Roxy, C. Gnanaseelan, *Indian Ocean Warming*. In *Assessment of Climate Change over the Indian Region*, 191-206. Springer publications, Singapore, 2020.
- [53] A. Yaduvanshi, A. Kulkarni, "Observed changes in extreme rain indices in semiarid and humid regions of Godavari basin, India: risks and opportunities", *Natural Hazards*, 103, 685-711, 2020.

## Design, Optimization and Simulation of a New Decoder for Reed Solomon and BCH Codes Using the New Syndromes Block

Mohamed Elghayaty<sup>\*1</sup>, Anas El Habti El Idrissi<sup>1</sup>, Omar Mouhib<sup>1</sup>, Azeddine Wahbi<sup>2</sup>, and Abdelkader Hadjoudja<sup>1</sup>

<sup>1</sup>Laboratory of Electrical System, Transmission of Information, Mechanics and Energetics, Ibn Tofail University, Kenitra, BP14000, Morocco

<sup>2</sup>Laboratory of Industrial Engineering, Data Processing and Logistic, Faculty of Sciences Ain Chock, University Hassan II, Casablanca, Morocco

### ARTICLE INFO

Article history:

Received: 13 October, 2022

Accepted: 23 December, 2022

Online: 24 January, 2023

Keywords:

RS codes

BCH codes

DVB-S and DVB-S2 transmission

Chains

Galois field

Syndrome block

Quartus, VHDL

### ABSTRACT

In this paper, a new syndrome block for Reed Solomon RS and BCH codes used respectively in digital Video broadcasting DVB-S and DVB-S2 has been presented in order to reduce the number of iterations compared to the existed block, which can be found in the literature, the new method is based on a factorization of the equation corresponds to the syndrome block, which allows us to conceive another circuit. However, this reduction can approximately attain 40%. First, we developed and concepted the design of the proposed algorithm. Second, we transformed the circuits on hardware description language VHDL and finally we generated and simulated the basic and proposed algorithms using Quartus software tools.

## 1. Introduction

The quality of a data digital transmission [1]-[3]. Largely depends on the number of errors introduced via the transmission channel. Error control by coding technique is important. Indeed, this technique called "channel coding"[4]-[6], permit both detection and correction of possible transmission errors by using error-correcting codes such as RS codes (Reed-Solomon.) [7] [8] BCH (Bose, Ray-Chaudhuri and Hocquenghem) [9], [10] and LDPC (Low-Density -Parity-Check) [11].

However, the "channel coding" technique [12], [13] uses a very complex decoding mechanism requiring a very large number of logic gates, which influences the response time.

The main aim of this work is to develop, concept and simulate a new architecture for RS and BCH codes in order to reduce the number of iterations in the syndrome block using a new method based on the factorization technic (factorization method: we develop and factorize the equation corresponds to the syndrome block, the basic circuit is transformed into a new circuit which the

inputs are parallel). Other points are noted like: a summary of Reed Solomon codes is furnished in chapter 2. chapter 3 talks about the proposed algorithm that uses a new syndrome Block .Finally comparison of the basic and the proposed circuits for various RS codes is presented in chapter 4, ended with a conclusion

## 2. Reed Solomon Code

The RS (255,239) code [14] has length  $n = 2^8 - 1 = 255$  so  $m = 8$ , which imply that the Galois field contains 256 symbols ( $m = 8$ ), where the polynomial of an element in the Galois field can be represented as:

$$a_7 x^7 + a_6 x^6 + a_5 x^5 + a_4 x^4 + a_3 x^3 + a_2 x^2 + a_1 x^1 + a_0 x^0 \quad (1)$$

The symbols are 8 bits. It is thus constructed from the **Galois Field** GF ( $2^8$ ) [15]. The control symbols is  $N-K=16$ ,  $t=N-K=2t=8$  symbols correctable by N-bit words. The correction power therefore corresponds to a maximum of 64 bits since each symbol is on 8 bits.

The efficiency of this code is given by is:

<sup>\*</sup>Corresponding Author: Mohamed Elghayaty, melghayaty@gmail.com

$$R=N/K= 239/255= 0.937 \quad (2)$$

This allows us to construct the elements of Galois fields GF (255).

The symbols used in RS codes are:

- $t$ =corrected errors number.
- $n$ = total number of symbols
- $K$ = symbols of message.
- $(n-k)$  = detected errors number
- $t$ = detected errors number

$$\text{We use } \alpha^8 = \alpha^7 + \alpha^4 + \alpha^3 + \alpha^2 + 1 \text{ to be able to code the whole element: } \alpha^8 = \alpha^7 + \alpha^4 + \alpha^3 + \alpha^2 + 1 \quad (4)$$

The  $\alpha^i$  for  $i$  ranging from [9; 254] can be obtained from the multiplication rule:

$$\alpha^{i+1} = \alpha \alpha^i \quad (5)$$

#### 4. Proposed of a new architecture for Syndrome Block

The proposed algorithm [18][19] of the Reed-Solomon code RS (255, 239) used in DVB-T has 86 iterations, while 256 iterations using the existed method .This algorithm is based on the new syndrome block to reduce the number of iterations with a percentage which can reach 40% compared to the existing algorithm.

*Basic syndrome computation block*

a) *Case of the basic circuit for RS (15,11)*

The basic syndrome computation block for RS (15, 11) is expressed by the equation 6.

$$S_i = R(\alpha^i) = r_{14}(\alpha^i)^{14} + r_{13}(\alpha^i)^{13} + \dots r_1(\alpha^i) + r_0 \quad (6)$$

In the equation 6, the circuit corresponding shown in the figure2:

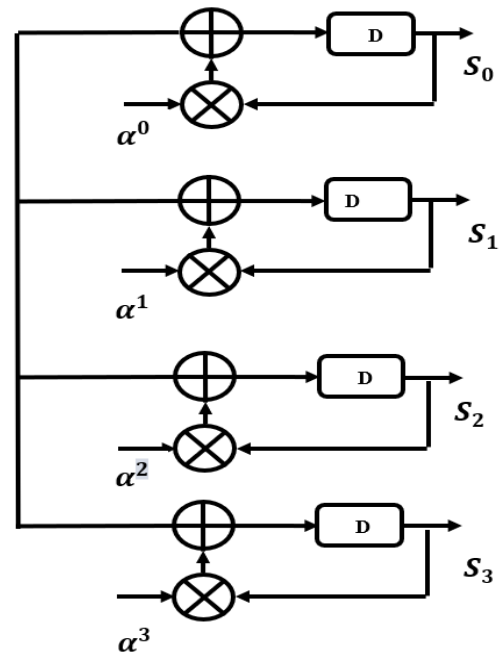


Figure 2: Basic syndrome block for RS (15, 11)

b) *Case of the basic circuit for RS (63,53)*

The basic syndrome computation block for RS (63, 53) is expressed by the equation 7.

$$S_i = R(\alpha^i) = r_{62}(\alpha^i)^{62} + r_{61}(\alpha^i)^{61} + \dots r_1(\alpha^i) + r_0 \quad (7)$$

In the equation 7, the circuit corresponding shown in the figure 3:

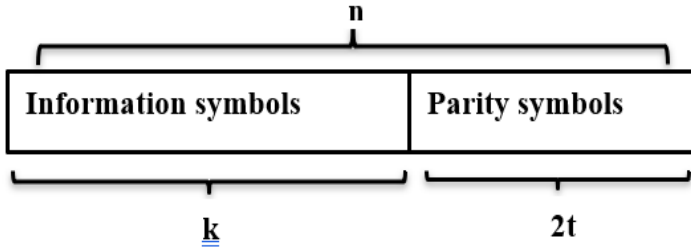


Figure 1: Reed Solomon code word structure

### 3. Galois Field (GF)

#### 3.1. Galois Field Properties

The principal properties of a Galois field [16] are:

- Two operations characterize the Galois Field: addition and multiplication.
- The result of addition or multiplication of Galois field elements allows us an element in the same field.
- For each element  $m$  in the field, “zero” is the Identity of addition, such that  $m + 0 = m$ .
- For each element  $m$  in the field, “one” is the Identity of multiplication, such that  $m * 1 = m$ .
- For each element  $m$  in the Galois field,  $n$  is an inverse of addition element such that  $m+n = 0$ .
- For each element  $m \neq 0$  in the Galois field,  $n^{-1}$  is an inverse of multiplication such that  $n*n^{-1}=1$ .
- Addition and multiplication operations should verify the laws of commutative, associative and distributive.

#### Galois Field GF (2<sup>m</sup>)

Knowing that Galois field [17] can be considered a general case to Binary Field. We hypothesize that we want to generate a finite field GF (q) where q a prime number.

For the Galois field GF (2<sup>8</sup>) = 256 symbols composed of 8 bits.

$$GF (2^8) = (0, \alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4 \dots \alpha_{254})$$

The corresponding primitive polynomial is:

$$P(x) = x^8 + x^7 + x^4 + x^3 + x^2 + 1 \quad (3)$$

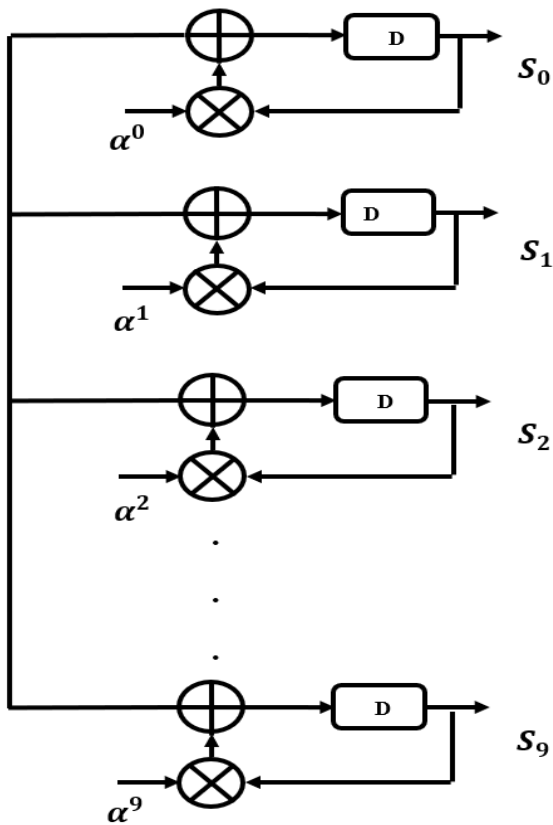


Figure 3: Basic Syndrome block for RS (63, 53)

The proposed Syndrome Computation Block

a) Case of the proposed circuit for RS (15, 11)

Using both equations 7 and 8, we can calculate all coefficients of syndrome block polynomial [20]:

$$S_i = R(\alpha^i) = r_{14}(\alpha^i)^{14} + r_{13}(\alpha^i)^{13} + \dots + r_1(\alpha^i) + r_0 \quad (7)$$

Where  $i = 1, 2, 3, \dots, 2t$ .

The proposed Syndrome computation Block calculated by this equation:

$$S_i = R(\alpha^i) = ((\dots (r_{14}(\alpha^i)^2 + r_{13}(\alpha^i)^1 + r_{12}(\alpha^i)^3 + r_{11}(\alpha^i)^2 + r_{10}(\alpha^i)^1 + r_9(\alpha^i)^3 + \dots + r_2(\alpha^i)^2 + r_1(\alpha^i) + r_0)) \quad (8)$$

The first clock, the mot received in parallel is  $(r_{14}, r_{13}, r_{12})$ .

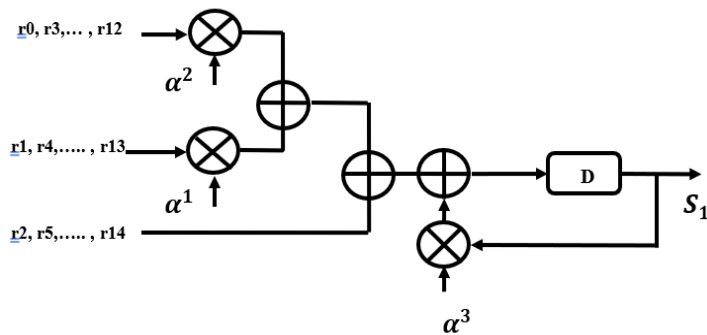


Figure 4: Proposed syndrome block for RS (15, 11)

b) Case of the proposed circuit for RS (63, 53)

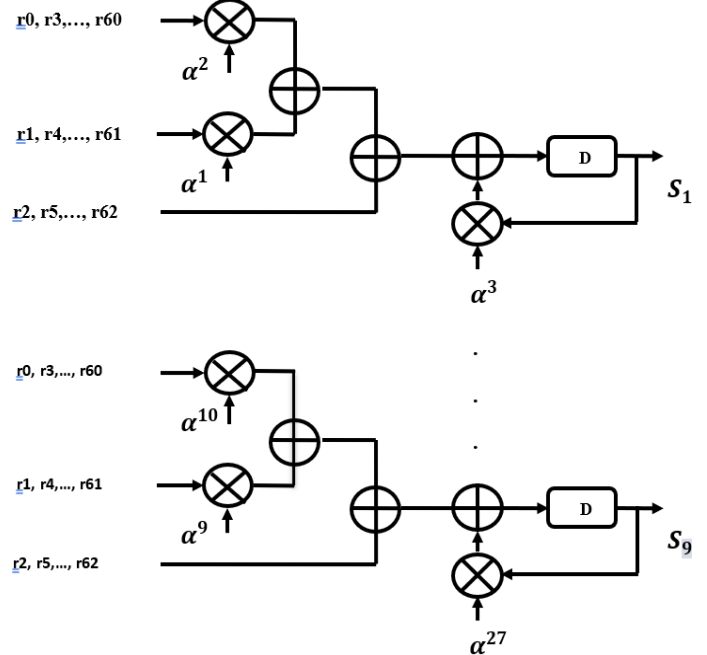


Figure 5: Proposed syndrome block for RS (63, 53)

c) Case of the Proposed circuit for RS (255,239)

For the case of the RS (255,239) we have:  $n-k = 2t = 16$  syndromes. For calculate of the example the syndrome  $S_1$  we have the circuit:

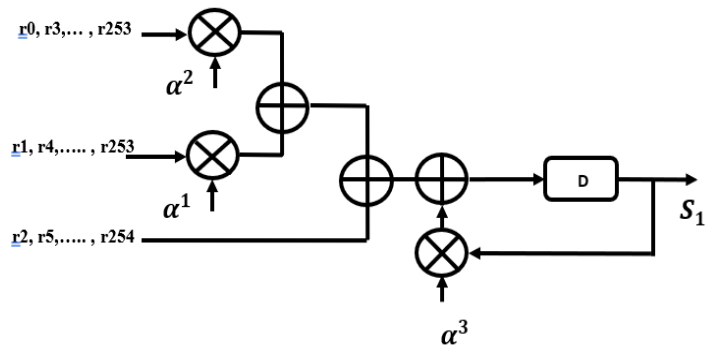


Figure 6: Proposed syndrome block for RS (255,239)

We generally use the following circuit:

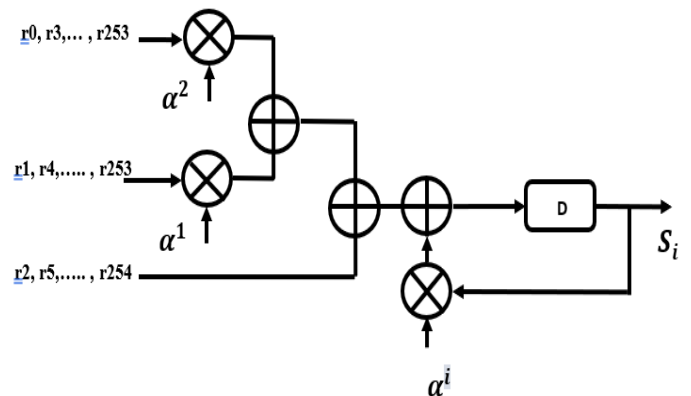


Figure 7: Proposed syndrome block for RS (255,239)



Table 1: Comparison of circuits and performance analysis

Code N°	Code Name	Number of iterations for the basic circuit (Nb)	Number of iterations for the proposed circuit (Nm)	Number of gained iterations
1	RS (15, 11)	16	6	10
2	RS (63, 55)	64	22	42
3	RS (255, 239)	256	86	170
4	RS (1023, 1019)	1024	342	682
5	RS (3240, 3070)	3241	1081	2159
6	RS (4095, 4091)	4096	1366	2730
...	.....	....	.....	.....
n	RS (n, k)	N <sub>b</sub> +1	(N <sub>m</sub> /3) + 1	N <sub>b</sub> - N <sub>m</sub>

*Comparison of Circuits*

For the table 1 shows the number for the basic, the proposed and the gained iterations for different Reed Solomon codes:

In the table 1 the Reed Solomon RS (255, 239) code used 256 iterations for the basic syndrome block, while just 86 iterations for the modified method. This algorithm use the new syndromes blocks to reduce the number of iterations. This method also reduces energy consumption with apercentage that can reach 33% compared to the existing algorithm.

*Performance analysis*

Proving the performance of the proposed algorithm, we fulfill an important number of checks in context of syndrome block for different RS code; different parallel syndrome block is tested. The Simulation result of RS (15, 11) is shown in the figure 9.

The figure 8 represents the different RS codes of the parallel syndrome block.

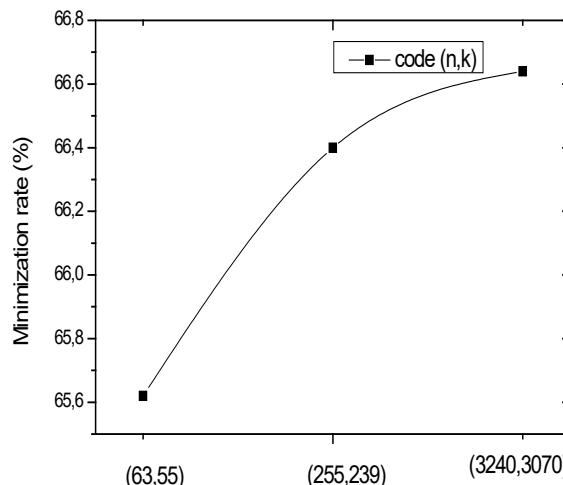


Figure 8: Evolution of the minimization rate the proposed circuit for the different RS codes.

**5. Simulation results**

The simulation of the basic and proposed syndrome block using the hardware description language VHDL [21] for the RS and BCH decoders are presented in this party.

*Simulation the proposed circuit of RS (15, 11)*

The Simulation result of the modified RS (15, 11) is shown in the figure 10.

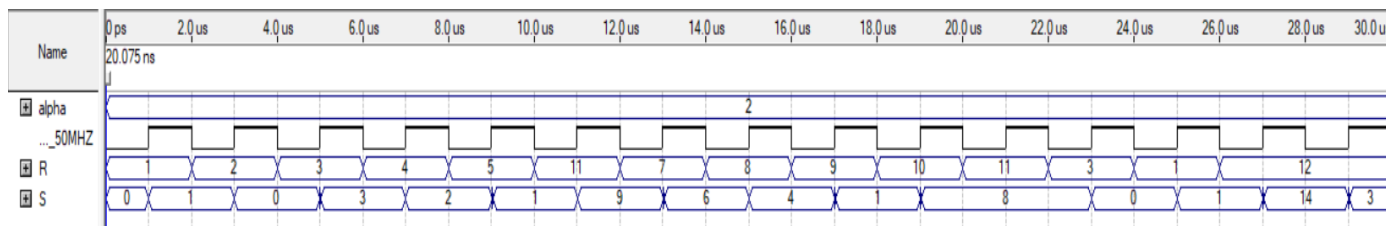


Figure 9: Simulation result of the basic decoder (15, 11).

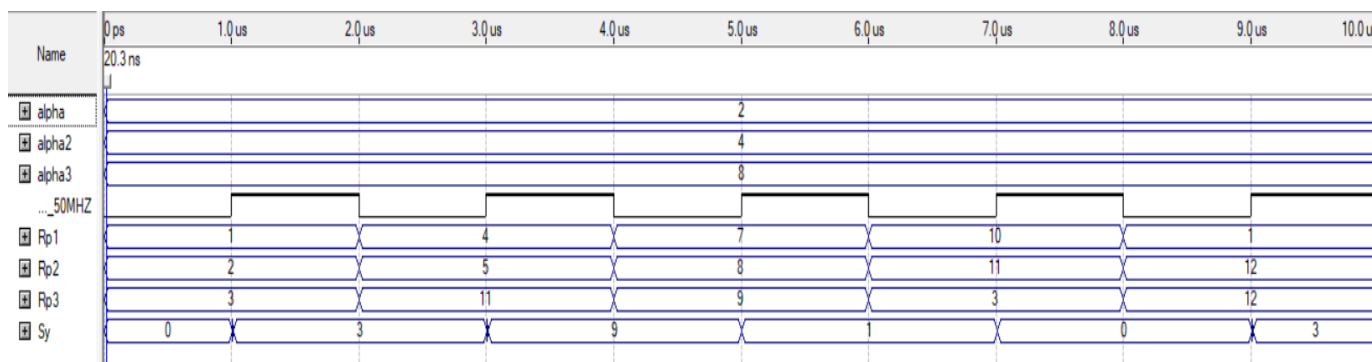


Figure 10: Simulation result of the modified decoder RS (15, 11)

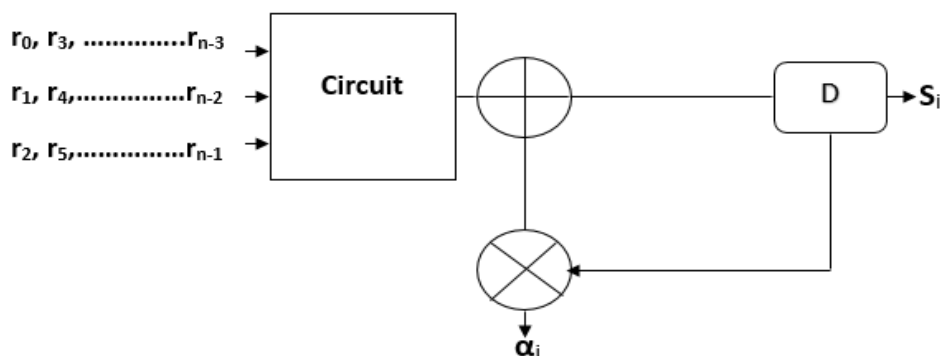


Figure11: Block diagram of Syndrome Block

*Simulation the basic circuit of RS (15, 11)*

The simulation of the developed and basic syndrome bloc is presented in this part for the RS and BCH decoders. Thus, simulation results on the tested scenario show that the proposed system is very effective and achieves high performance in the minimization rate.

**6. FPGA implementation**

Implementation of decoder algorithms for Reed-Solomon codes [22], [23] can be considered as a problematic cases on account of the very large amount of used electronic elements in order to implement the new algorithm on FPGA card to discuss how to save the hardware resources [24], [25]. In this paper a new hardware model of the Syndrome Block has been conceived and developed using the programming Language (VHDL) and implemented using Xilinx Synthesis Tool. The circuit scheme of the implemented program is shown in Figure11.

The proposed Syndrome Block consists of a global 'Clk' and Three Parallel Inputs initiate the calculating Syndrome Block process, the 'result' can be obtained immediately after entering inputs.

*Syndrome Block*

The calculation of the syndrome block furnishes us two results: 1- all syndrome polynomial coefficients are equal to zero, in this case we stop the rest of the decoder process because the received code word is correct, 2- if one of polynomial coefficients is different to zero, the code word is erroneous, so we continue the process of the decoder. We need 2t basic scheme as defined in Fig.12. Where  $1 \leq i \leq 2t$ , or for each Syndrome  $S_i$ , n iterations are needed to calculate the polynomial coefficients.

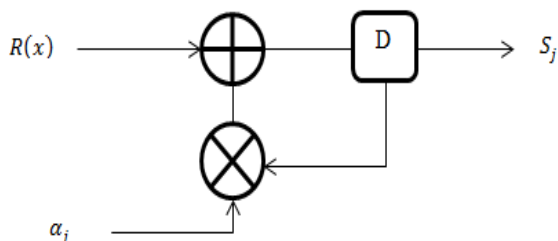


Figure12: Basic syndrome calculator cell

*Test procedure for RS (15, 11)*

The proposed algorithm has been implemented on a FPGA Card using Xilinx Spartan 3E-500 to verify the test setup which presented in figure13.



Figure 13: Value of Syndrome block for RS (15, 11) codes

For the case of RS (15, 11), we have four coefficients of syndrome Block ( $S_0, S_1, S_2, S_3$ ). In Fig.13, the value is equal to 15 ( $S_0=15$ ) in decimal, (1111) in binary, so we can get the same result with only 5 iterations in comparison with the basic circuit.

*The code specified for DVB-T*

The Digital Video Broadcasting-T standard defines RS (255, 239, 8) code, a main version is proposed to generate (204, 188, 8) code, this code contain 204 symbols, where 188 represent the symbols of message [8]. The Galois field of RS (255, 239) code has 256 symbols ( $m=8$ ) so we can represent the polynomial of a field element as:

$$a_7 x^7 + a_6 x^6 + a_5 x^5 + a_4 x^4 + a_3 x^3 + a_2 x^2 + a_1 x^1 + a_0 \quad (9)$$

Where the polynomial generator for  $t = 8$ , can be presented as:

$$P(x) = x^8 + x^4 + x^3 + x^2 + 1 \quad (10)$$

For the case of RS (255, 239) used in Digital Video Broadcasting-T standard, the decoder detects  $2t=16$  errors and corrects  $t=8$  errors.

*Test procedure for RS (255, 239)*

The proposed algorithm has been implemented on a FPGA Card using Xilinx Spartan 3E-500 to verify the test setup which presented in figure14.

For the case of RS (255, 239), we have four coefficients of syndrome Block (S0, S1, S2, S3). In Figure 14 the value is equal to 186 (S0 = 186) in decimal, (10111010) in binary, so we can get the same result with only 5 iterations in comparison with the basic circuit.

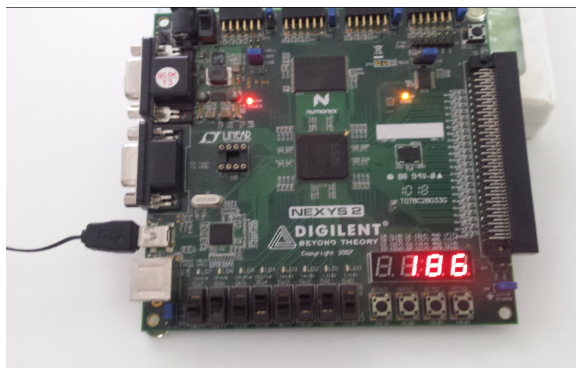


Figure 14: Value of Syndrome block for RS (255, 239) codes

## 7. Conclusion

A recent algorithm of syndrome block for Reed-Solomon RS and BCH codes has been presented in this paper. This algorithm presents a new syndrome computation block with a view to minimize the number of iterations. The proposed algorithm has been generated, simulated, implemented on the FPGA card and compared to the existed one to demonstrate the difference between the two circuits and the number of reduced iterations, the comparison between circuits in table 1 proves that the RS code (255, 239) has 256 iterations using the modified method while, 86 iterations using the basic method.

## Conflict of Interest

The authors declare no conflict of interest.

## Acknowledgments

This work was supported by the Laboratory of Electrical System, Transmission of Information, Mechanics and Energetics, Faculty of Sciences, Ibn Tofail University Kenitra, Morocco.

## References

- [1] R. Huynh, N. Ge, H. Yang, "A Low Power Error Detection in the Syndrome Calculator Block for Reed-Solomon Codes: RS(204,188)", *Tsinghua Science and Technology*, **14**(4), 474 - 477, 2009, doi:10.1016/S1007-0214(09)70105-1.
- [2] Y.J. Tang, X. Zhang, "Fast En/Decoding of Reed-Solomon Codes for Failure Recovery", *IEEE Transactions on Computers*, **71**(3), 724 - 735, 2022, doi:10.1109/TC.2021.3060701.
- [3] V. Torres, J. Valls, M.J. Canet, F. García-Herrero, "Soft-decision low-complexity chase decoders for the RS(255,239) code", *Electronics (Switzerland)*, **8**(1), 1 - 13, 2019, doi:10.3390/electronics8010010.
- [4] R.T. Chien, "Cyclic Decoding Procedures for Codes », *IEEE Transactions on Information Theory*, **10**(4), 357-362, 1965.
- [5] P.D. Surkar, S.D. Ninawe, "VLSI Design of Syndrome Computation Block for RS ( 255 , 239 ) Code", 5248 - 5254, 2016, doi:10.15680/IJIRSET.2016.0504130.
- [6] D.S. Reay, T.C. Green, B.W. Williams, "Field programmable gate array implementation of a neural network accelerator ", *IEE Colloquium (Digest)*, (61), 1994.
- [7] R. Martinek, J. Zidek, "The implementation of channel coding into the digital transmission chain consisting of VSG PXI-5670 - VSA PXI-5661 ", *Przeglad Elektrotechniczny*, **89**(7), 64 - 68, 2013.

- [8] W. Ji, W. Zhang, X. Peng, Y. Liu, "High-efficient Reed-Solomon decoder design using recursive Berlekamp-Massey architecture ", *IET Communications*, **10**(4), 381-386, 2016, doi:10.1049/iet-com.2015.0500.
- [9] Y.H. U, M.R. Hiremath, "Implementation of BCH Code (n, k) Encoder and Decoder for Multiple Error Correction Control ", *International Journal of Computer Science and Mobile Applications*, **2**(5), 45-54, 2014.
- [10] S.S. Sonawane Vaishali Baste, "Implementation of RS-CC Encoder & Decoder using MATLAB", *IJSTE-International Journal of Science Technology & Engineering*, **5**(7), 22-30, 2019.
- [11] C. Sahana, V. Anandi, "INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY ERROR DETECTION USING BINARY BCH (255, 215, 5) CODES Sahana C\*, V Anandi ", **4**(6), 1113-1118, 2015.
- [12] P. Trifonov, V. Miloslavskaya, C. Chen, Y. Wang, "Fast encoding of polar codes with reed-solomon kernel", *IEEE Transactions on Communications*, **64**(7), 2746-2753, 2016, doi:10.1109/TCOMM.2016.2576448.
- [13] M. Elghayyaty, O. Mouhib, A. Wahbi, A.A. Barakate, A.E.H. El Drissi, L. Hlou, A. Hadjoudja, "Performance comparison of new designs of chien search and syndrome blocks for BCH and Reed Solomon codes ", *International Journal of Communication Networks and Information Security*, **12**(2), 235-241, 2020, doi:10.17762/ijcnis.v12i2.4562.
- [14] M. Elghayyaty, A. Wahbi, A. El Habti El Drissi, O. Mouhib, L. Hlou, A. Hadjoudja, "Conception and Hardware Minimization of a New Chien Search Block for Reed Solomon Codes With Implementation on Fpga Card ", *ARNP Journal of Engineering and Applied Sciences*, **15**(11), 1248-1254, 2020.
- [15] D. Gunduz, "Source and Channel Coding for Wireless Networks ", (September), 2007.
- [16] J.L. Massey, "Step-by-step decoding of BCH codes ", *IEEE Transactions on Information Theory*, **11**(3), 3-8, 1965.
- [17] G.A. Hussain, L. Audah, "BCH codes in UPMC: A new contender candidate for 5G communication systems ", *Bulletin of Electrical Engineering and Informatics*, **10**(2), 904-910, 2021, doi:10.11591/eei.v10i2.2080.
- [18] E. Costa, S.V. Fedorenko, P.V. Trifonov, "On computing the syndrome polynomial in Reed-Solomon decoder ", in *European Transactions on Telecommunications*, 2004, doi:10.1002/ett.982.
- [19] Z.Y. Lam, W.L. Pang, C.P. Ooi, S.K. Wong, K.Y. Chan, "VHDL modelling of Reed Solomon decoder ", *Research Journal of Applied Sciences, Engineering and Technology*, **4**(23), 5193-5200, 2012.
- [20] M. Prashanthi, P. Samundiswary, "An Area Efficient (31, 16) BCH Decoder for Three Errors ", *International Journal of Engineering Trends and Technology*, **10**(13), 616-620, 2014, doi:10.14445/22315381/ijett-v10p323.
- [21] Z. Gao, L. Zhang, Y. Cheng, K. Guo, A. Ullah, P. Reviriego, "Design of FPGA-Implemented Reed-Solomon Erasure Code (RS-EC) Decoders with Fault Detection and Location on User Memory ", *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, **29**(6), 1073 - 1082, 2021, doi:10.1109/TVLSI.2021.3066804.
- [22] J. Samanta, J. Bhaumik, S. Barman, "FPGA based area efficient RS(23, 17) codec ", *Microsystem Technologies*, **23**(3), 639 - 650, 2017, doi:10.1007/s00542-016-3058-1.
- [23] C. Engineering, M. Prashanthi, P. Samundiswary, M. Tech, "An Enhanced (15, 5) BCH Decoder Using VHDL ", *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Energy*, **2014**(11), 2014.
- [24] K.M.M. Chennaiah, K. Prasadbabu, S. Ahmedbasha, "IMPLEMENTATION OF BCH LFSR ENCODER DECODER ", **5**(1), 21 - 30, 2017.
- [25] P. Mathew, L. Augustine, S. G. T. Devis, "Hardware Implementation of (63, 51) Bch Encoder and Decoder for Wban Using LFSR and BMA ", *International Journal on Information Theory*, **3**(3), 1 - 11, 2014, doi:10.5121/ijit.2014.3301.

## Nonlinear Model Predictive Control of Rover Robotics System

Serdar Kalaycioglu\*, Anton de Ruiter

Department of Aerospace Engineering, Toronto Metropolitan University, Toronto, M5B 2K3, Canada

### ARTICLE INFO

Article history:

Received: 30 September, 2022

Accepted: 23 December, 2022

Online: 24 January, 2023

Keywords:

NMPC

Optimal Control

Multi Rover Control

### ABSTRACT

The paper presents two robust and efficient control algorithms based on (i) Optimal Control Allocation (OCA) and (ii) Nonlinear Model Predictive Control (NMPC). The robotics system consists of two rovers with mecanum wheels and mounted two 7-DOF arms carrying a common load. The overall system is an underdetermined one with non-holonomic constraints. The developed control algorithms focus on providing an optimal solution to the wheel and joint torque saturation problem, which is typically encountered while manipulating a large and heavy payload. The first control algorithm based on OCA minimizes a quadratic cost function consisting of robot joint and rover wheel torques, contact forces, and moments using only the current state values and the system dynamics. It is computationally very efficient. The NMPC algorithm minimizes a quadratic cost function which not only includes the current states but also the future state estimates, and the control inputs over a specified prediction horizon. The system consisting of multi-rover with a dual arm is highly non-linear. The linear MPC technique on which most of the previous studies relied is not adequate. On the other hand, the computational difficulties of a generic NMPC algorithm is remarkably high. In this paper, an elegant, discretized technique with exact realization is implemented to take into account the full non-linear model and yet provide a simple real-time solution satisfying a minimum performance index subject to constraints. The results show that the developed control algorithms OCA and NMPC work efficiently, and the minimum the contact moments and forces, and the joint torques are realized while two arms carry a common load and successfully track a reference end-effector trajectory. The results also indicate that although NMPC algorithm is computationally more involved, it provides superior results in reducing joint and wheel torques as well as contact moments and forces.

## 1. Introduction

This paper is an extension of the work originally presented in IEMTRONICS [1]. The Optimal Control Allocation algorithm (OCA) presented in the original work is further extended to accommodate a Nonlinear Model Predictive Control technique to increase performance of the approach.

There has been a significant interest in exploring complex environments using mobile rovers. Such rovers are commonly used in space exploration, construction, mining, and military.

Especially, there has been a considerable amount of interest in Space Robotics Exploration missions in the last two decades. Similar to on-orbit robotics missions (e.g., servicing, assembly,

and manufacturing), the future planetary exploration missions will also include tasks such as assembly of large space structures using multiple coordinating rovers and the rover-mounted robotics manipulators. Recently the Moon and Mars rover missions are the main target of various space agencies including NASA, Canadian Space Agency, ESA, JAXA, etc. Most of these space agencies in collaboration with space industries and research centers are heavily focusing on innovative rover technologies and designs. Autonomous rover motion control capability has been identified as one of the critical and enabling technology requirement for such systems. Although, there is a significant amount of research studies in the fields of control of single rover trajectory and force control of fixed-based arms, there are still major research challenges in the areas of load sharing multi-rovers and arms, particularly, real-time force and motion control when they are carrying a common load.

\* Corresponding Author: Serdar Kalaycioglu, TMU, Department of Aerospace Engineering, Toronto, Canada, email: [skalay@torontomu.ca](mailto:skalay@torontomu.ca)

The initial technological challenges that involved designing a mobile rover were related to its mechanics. These included the development of dynamic control systems and collision free trajectories.

In order to develop effective control systems for mobile rovers, a team led by Neculescu [2,3] studied the free and contact motion of the vehicles. They also developed methods to generate collision free trajectories and perform force control.

Motion control of rovers with nonholonomic constraints were studied using differential wheeled rovers in [4,5] These constraints exist if the constraints cannot be expressed in the form of time derivatives of a function consists of the generalized coordinates.

There have been extensive studies in control of systems with non-holonomic constraints. However, most of the cases, kinematic control is typically achieved by ignoring the dynamics when dealing with systems with non-holonomic constraints [6]. However, it has been shown that a mechanical system with these constraints were controlled in spite of its structure [7]. In addition, it has been shown that a non-holonomic system cannot be brought to a single equilibrium with a smooth time-invariant feedback [8].

In a study conducted in Kalaycioglu [9], a control technique with optimal force distribution for multiple robotic manipulators was demonstrated. However, it only involved two cooperating arms and did not include rovers.

The use of a Model Predictive Control (MPC) framework facilitates the optimization of a given performance index. It also allows for the analysis of the system constraints and dynamics [10–15]. One of the most challenging aspects of implementing a robust model of (MPC) is dealing with the various uncertainties that can impact its performance [16]. Due to the characteristics of the model's receding horizon, standard MPC can provide an adequate level of robustness [17].

Unfortunately, the literature has shown that standard MPC cannot provide an adequate performance in complex robotics systems [18]. To address this issue, various research studies have been conducted to develop novel MPC methods that can provide a robust and stable performance [19–23].

The scope and capabilities of Non-linear Model Predictive Control (NMPC) have significantly improved over the past few years. Due to the increasing number of tools that can be used to implement this type of model, the performance of this algorithm has been greatly improved. Some of these include the ability to perform fast gradient use and input parameterisation [24–27]. The application of NMPS for free-floating space manipulator are provided in [28–31].

The mechanics of wheeled locomotion have also attracted a lot of attention [32–37]. A number of studies have been conducted on the dynamics and kinematics of the mecanum wheel (a subcategory of omnidirectional wheel) [38–43].

There has been a significant amount of research on the various aspects of wheeled locomotion, but it is still not yet feasible to fully understand the mechanisms involved in the movement control of multiple rovers and mounted arms. For instance, the development of systems with multi- rovers with dual manipulators that can

perform real-time trajectories while manipulating a common load is still in its early stages.

This paper presents two robust and efficient control algorithms based on (i) Optimal Control Allocation (OCA) and (ii) Nonlinear Model Predictive Control (NMPC) for a rover robotics system with mecanum wheels when the two 7-DOF arms operating a common load. The system is an underdetermined one subject to non-holonomic constraints. The control algorithms focus on providing an optimal solution to the wheel and joint torque saturation problem, which is typically encountered while manipulating a large and heavy payload.

The first control algorithm based on OCA minimizes a quadratic cost function (a performance index) consisting of robot joint and rover wheel torques, contact forces, and moments using only the current state values and the system dynamics. It is computationally very efficient. The NMPC algorithm minimizes a quadratic cost function which not only includes the current states but also the future state estimates, and the control inputs over a specified prediction horizon.

The literature on the application of MPC for robotics is mainly focused on linear models. However, the multi-rover dual arm coordinating system is highly non-linear and MPC lacks robust applications in this area. In this paper, we present a novel NMPC discretized technique that incorporates the full non-linear characteristics of the multi-rover dual arm system.

This paper consists of four sections. The first section provides the mathematical formulations such as the kinematics and dynamics models of the total system including two  $n$ -degree redundant manipulators, two rovers and a common load. The second section presents two novel control algorithms based on optimal control allocation (OCA) and non-linear model predictive control (NMPC) which are formulated to minimize the wheel moments, the joint torques, and contact moments/forces. The third section provides the simulation results and discussion, and the fourth section provides some concluding remarks and recommendations for future work.

## 2. Theoretical Formulations

### 2.1. The Rover Robotics System

The system includes two mobile rovers with four mecanum wheels and two  $n$ -DOF redundant arms attached on the two rovers carrying a common load. Figure 1 shows an example of such a system with two rovers and two  $n$ -degree arms.

Table 1 contains the rover and robotics parameters utilized in the computer simulations.. The rotation angle  $\psi_i$  and the position vector  $\tilde{\mathbf{R}}_{ci}$ , provide the pose of the center of mass  $C_i$  of the  $i^{\text{th}}$  rover-in the inertial coordinate system, X, Y, Z. The coordinate axes  $x_{ci}$ ,  $y_{ci}$ ,  $z_{ci}$  attached to point  $C_i$  are obtained via a rotation around Z-axis with an angle of  $\psi_i$ .

The masses associated with the rovers and the wheels are given as  $m_{ci}$  and  $m_{wij}$ , respectively for the  $i^{\text{th}}$  rover and the  $j^{\text{th}}$  wheel,  $j=1...4$  and  $i=1,2$  for each rover. The distances between the wheel centers along the  $y_{ci}$  and  $x_{ci}$ -axes are denoted by  $2a$  and  $2b$ , respectively.

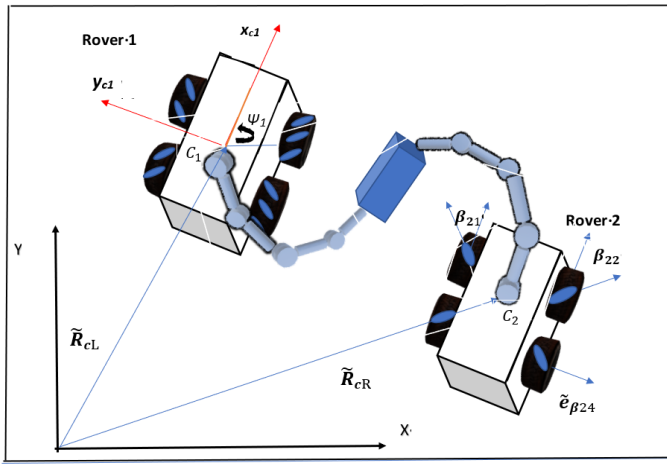


Figure 1: Description of the rover robotics system

The wheels have a radius of  $s$  and the angle of rotation, and the angular rate are denoted as  $\phi_{ij}$  and  $\omega_{ij}$ , respectively. The rollers are attached to the outer rims of the mecanum wheels as illustrated in Figure 1. The angle  $\beta_{ij}$  is defined as the angle between the axis of rotation of the roller and the  $x_{ci}$  of the  $j^{\text{th}}$  wheel of the  $i^{\text{th}}$  rover.

## 2.2. Model of Kinematics

$\tilde{V}_{mij}$ , the velocity vector of the center of the  $j^{\text{th}}$  wheel of the  $i^{\text{th}}$  rover can be determined by the following relationship:

$$\tilde{V}_{mij} = \tilde{V}_{ci} + \tilde{\Omega}_{ci} \times \tilde{r}_{wij} \quad (1)$$

$$\tilde{\Omega}_{ci} = \dot{\psi}_i \tilde{e}_z \quad (2)$$

where  $\tilde{V}_{ci}$  is the velocity of the mass center of the rover,  $\tilde{\Omega}_{ci}$  is the angular velocity vector of the rover and  $\tilde{e}_z$  is a unit vector both along the  $z_{ci}$  axis while  $\tilde{r}_{wij}$  is the position vector from the rover's mass center to the wheel center.

The velocity vector  $\tilde{V}_{pij}$ , representing the velocity of a point  $P$  located at the roller center can be expressed as

$$\tilde{V}_{pij} = \tilde{V}_{mij} + \tilde{\omega}_{ij} \times \tilde{\rho}_{ij} \quad (3)$$

where  $\tilde{\rho}_{ij}$  is the position vector from the wheel's center to the point  $P$ , the roller center.

If the rollers do not slip,  $\tilde{V}_{pij}$  does not have a component in the direction of the axis of roller rotation  $\tilde{e}_{\beta ij}$ , and can be expressed as

$$\tilde{V}_{pij} \cdot \tilde{e}_{\beta ij} = 0 \quad (4)$$

where  $\tilde{e}_{\beta ij}$  is a unit vector along the roller's axis of rotation. After carrying out some algebraic manipulations using (3) and (4), one can write the following expressions:

$$\begin{aligned} \tilde{V}_{mij} \cdot \tilde{e}_{\beta ij} + (\tilde{\omega}_{ij} \times \tilde{\rho}_{ij}) \cdot \tilde{e}_{\beta ij} &= 0 \\ (\tilde{\omega}_{ij} \times \tilde{\rho}_{ij}) &= -\tilde{\omega}_{ij} s \tilde{e}_{xi} \end{aligned}$$

$$\tilde{V}_{mij} \cdot \tilde{e}_{\beta ij} = \tilde{\omega}_{ij} s (\tilde{e}_{xi} \cdot \tilde{e}_{\beta ij}) \quad (5)$$

where  $s$  is the radius of the wheel and  $\tilde{e}_{xi}$  is a unit vector in the direction of the  $x_{ci}$  axis.

Furthermore, rewriting the equations of constraints by utilizing (1) and (5), one can obtain the following relationships:

$$\begin{aligned} \tilde{V}_{ci} \cdot \tilde{e}_{\beta ij} + (\tilde{\Omega}_{ci} \times \tilde{r}_{wij}) \cdot \tilde{e}_{\beta ij} &= \tilde{\omega}_{ij} s (\tilde{e}_{xi} \cdot \tilde{e}_{\beta ij}) \\ \tilde{V}_{ci} \cdot \tilde{e}_{\beta ij} + (\tilde{r}_{wij} \times \tilde{e}_{\beta ij}) \cdot \tilde{\Omega}_{ci} &= \tilde{\omega}_{ij} s \cos(\beta_{ij}) \end{aligned} \quad (6)$$

where  $\beta_{ij}$  is defined as the angle between the two unit vectors  $\tilde{e}_{xi}$  and  $\tilde{e}_{\beta ij}$

$$\begin{aligned} e_{\beta i1}^T &= [\cos(\beta_{i1}), -\sin(\beta_{i1}), 0] \\ e_{\beta i2}^T &= [\cos(\beta_{i2}), \sin(\beta_{i2}), 0] \\ e_{\beta i3}^T &= [\cos(\beta_{i3}), \sin(\beta_{i3}), 0] \\ e_{\beta i4}^T &= [\cos(\beta_{i4}), -\sin(\beta_{i4}), 0] \\ \tilde{r}_{wi1}^T &= [a, b, 0] \\ \tilde{r}_{wi2}^T &= [a, -b, 0] \\ \tilde{r}_{wi3}^T &= [-a, b, 0] \\ \tilde{r}_{wi4}^T &= [-a, -b, 0] \end{aligned} \quad (7)$$

One can obtain the following expressions by plugging (7) into (6) and substituting  $45^\circ$  for  $\beta_{ij}$ :

$$\begin{aligned} \tilde{V}_{ci} &= \begin{bmatrix} V_{cix} \\ V_{ciy} \\ V_{ciz} \end{bmatrix} = \begin{bmatrix} s(\omega_{i1} + \omega_{i2})/2 \\ s(\omega_{i3} - \omega_{i1})/2 \\ 0 \end{bmatrix} \\ \tilde{\Omega}_{ci} &= \begin{bmatrix} \Omega_{cix} \\ \Omega_{ciy} \\ \Omega_{ciz} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ s(\omega_{i3} - \omega_{i1})/(2(a+b)) \end{bmatrix} \\ \omega_{i4} &= \omega_{i1} + \omega_{i2} - \omega_{i3} \end{aligned} \quad (8)$$

The following rotational matrix represents the rotation between the inertial and the rover body axes:

$$\underline{\Psi}_{zi} = \begin{bmatrix} \cos\psi_i & -\sin\psi_i & 0 \\ \sin\psi_i & \cos\psi_i & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (9)$$

Homogeneous transformation matrix  $\underline{T}_f^g$  which transforms the coordinates between frame-g and frame-f on the robot arm can be obtained by Denavit-Hartenberg (D-H) convention as follows.

$$\underline{T}_f^g = \underline{A}_{f+1} \underline{A}_{f+2} \dots \underline{A}_{g-1} \underline{A}_g \quad f < g$$

$$\underline{A}_f = \begin{bmatrix} \cos \theta_{fi} & -\sin \theta_{fi} \cos \alpha_{fi} & \sin \theta_{fi} \sin \alpha_{fi} & a_{fi} \cos \theta_{fi} \\ \sin \theta_{fi} & \cos \theta_{fi} \cos \alpha_{fi} & -\cos \theta_{fi} \sin \alpha_{fi} & a_{fi} \sin \theta_{fi} \\ 0 & \sin \alpha_{fi} & \cos \alpha_{fi} & d_{fi} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (10)$$

where  $\theta_{fi}, \alpha_{fi}, d_{fi}, a_{fi}$  are the parameters related to joint-f and link-f on the  $i^{\text{th}}$  arm, namely  $d_{fi}$  is the offset,  $a_{fi}$  is the  $f^{\text{th}}$  link-length,  $\theta_{fi}$  is the joint angle and  $\alpha_{fi}$  is the twist as defined in DH convention.

The following expressions can be used to obtain the Jacobian matrices and the first-time derivatives of these matrices associated with the rover's center and any arbitrary point-k on the arm:

$$\left(\underline{J}_c^k\right)_i = \begin{bmatrix} \tilde{e}_z & \tilde{e}_1 & \dots & \tilde{e}_7 \\ \tilde{e}_z \times \tilde{r}_{ck} & \tilde{e}_1 \times \tilde{r}_{1k} & \dots & \tilde{e}_7 \times \tilde{r}_{7k} \end{bmatrix} \quad (11)$$

$$\left(\underline{\dot{J}}_c^k\right)_i = \begin{bmatrix} \tilde{e}_z & \tilde{e}_1 & \dots & \tilde{e}_7 \\ \tilde{e}_z \times (\tilde{\Omega}_{ci} \times \tilde{r}_{ck}) & \tilde{e}_1 \times (\dot{\theta}_1 \tilde{e}_1 \times \tilde{r}_{1k}) & \dots & \tilde{e}_7 \times (\dot{\theta}_7 \tilde{e}_7 \times \tilde{r}_{7k}) \end{bmatrix} \quad (12)$$

Where  $\tilde{e}_i$  is the unit vector along the  $i^{\text{th}}$  joint rotation axis,  $\tilde{e}_z$  is the unit vector along  $\tilde{\Omega}_{ci}$ , and  $\tilde{r}_{ik}, \tilde{r}_{ck}$  are the position vectors from  $i^{\text{th}}$  joint and the rover's center to the point k, respectively.

The linear and angular velocities and accelerations of point k on the  $i^{\text{th}}$  rover arm can be calculated as follows:

$$\begin{bmatrix} \tilde{\Omega}_k \\ \tilde{V}_k \end{bmatrix}^i = \left(\underline{\dot{J}}_c^k\right)_i \begin{bmatrix} \tilde{\Omega}_{ci} \\ \dot{\theta}_i \end{bmatrix} + \begin{bmatrix} 0 \\ \tilde{V}_{ci} \end{bmatrix} \quad (13)$$

$$\begin{bmatrix} \dot{\tilde{\Omega}}_k \\ \dot{\tilde{V}}_k \end{bmatrix}^i = \left(\underline{\dot{J}}_c^k\right)_i \begin{bmatrix} \dot{\tilde{\Omega}}_{ci} \\ \ddot{\theta}_i \end{bmatrix} + \left(\underline{\dot{J}}_c^k\right)_i \begin{bmatrix} \tilde{\Omega}_{ci} \\ \dot{\theta}_i \end{bmatrix} + \begin{bmatrix} 0 \\ \dot{\tilde{V}}_{ci} \end{bmatrix} \quad (14)$$

where  $\begin{bmatrix} \tilde{\Omega}_k \\ \tilde{V}_k \end{bmatrix}^i$  is a vector consisting of the angular and linear velocity vectors of the point k on the  $i^{\text{th}}$  arm, respectively while  $\dot{\tilde{\theta}}_i^T = [\dot{\theta}_{i1}, \dot{\theta}_{i2}, \dot{\theta}_{i3}, \dots, \dot{\theta}_{in}]$  is a vector consists of the  $i^{\text{th}}$  rover-arm joint angular rates.

### 2.3. Model of Dynamics

The dynamics equations of motions of the system is derived using the Lagrangian formulation. The total kinetic energy  $T_t$  consists of two parts, the rotational and translational kinetic energies of the robotics arms and the rovers.

$$T_t = T_{tr} + T_{rt} \quad (15)$$

The angular and translational velocities of the rovers as well as that of the robot links' center of mass can be calculated using

(14) and (8). Then, the total kinetic energy of the system can be obtained using (15).

The dynamics equations of motion can be obtained using the following Lagrangian formulation:

$$\frac{d}{dt} \left( \frac{\partial T_t}{\partial \dot{q}_h} \right) - \frac{\partial T_t}{\partial q_h} = Q_h, \quad h = 1, \dots, 2m \quad (16)$$

where  $q_h$  and  $Q_h$  are the generalized coordinates and forces, respectively and

$$q^T = [\phi_{11}, \phi_{12}, \phi_{13}, \phi_{21}, \phi_{22}, \phi_{23}, \theta_{11}, \dots, \theta_{1n}, \theta_{21}, \dots, \theta_{2n}]$$

and  $m=(n+3)$ ,  $n$  represents the total number of degrees of freedom of the robotics arms.

Applying (16), the dynamics equations of motions for both rovers and the arms can be written in the following form:

$$\begin{bmatrix} G_{WL} & G_{WLR} & G_{W\theta L} & G_{W\theta R} \\ G_{WLR}^T & G_{WR} & G_{W\theta L} & G_{W\theta R} \\ G_{W\theta L}^T & G_{W\theta L} & G_{\theta L} & G_{\theta LR} \\ G_{W\theta R}^T & G_{W\theta R} & G_{\theta LR} & G_{\theta R} \end{bmatrix} \begin{bmatrix} \ddot{\Phi}_L \\ \ddot{\Phi}_R \\ \ddot{\theta}_L \\ \ddot{\theta}_R \end{bmatrix} + \begin{bmatrix} \tilde{c}_L \\ \tilde{c}_R \\ \tilde{c}_{\theta L} \\ \tilde{c}_{\theta R} \end{bmatrix} = \begin{bmatrix} \tilde{M}_L \\ \tilde{M}_R \\ \tilde{\tau}_{\theta L} \\ \tilde{\tau}_{\theta R} \end{bmatrix} \quad (17)$$

where  $G$  is the mass / inertia matrix (a positive definite matrix) and,  $\ddot{\Phi}_L, \ddot{\Phi}_R$  are the wheels' angular accelerations for the two rovers  $i=L$  and  $R$ , and  $\ddot{\theta}_L, \ddot{\theta}_R$  are the joint rotational accelerations for the two manipulators,  $i=L$  and  $R$ , respectively. The indices L and R are referred to the first and second rover and robotics arm, respectively.

The non-linear Coriolis and centrifugal terms are represented by  $\tilde{c}_{L^L}, \tilde{c}_R, \tilde{c}_{\theta L}$ , and  $\tilde{c}_{\theta R}$ . and  $\tilde{\tau}_{\theta L}, \tilde{\tau}_{\theta R}$  are the joint control torques for the two robot manipulators. Finally,  $\tilde{M}_L, \tilde{M}_R$  are the wheel control moments for the two rovers.

$\dot{\tilde{\Phi}}_i^T = [\omega_{i1}, \omega_{i2}, \omega_{i3}]$  includes the wheels angular rates of the  $i^{\text{th}}$  rover,  $i=L$  and  $R$  for two rovers. If there is no slip, the fourth wheel angular rate can be calculated using (8).

### 2.4. Optimal Control Allocation (OCA) Technique

The robotics system composed of two rovers and two redundant arms is an undetermined because of the excessive number of sensors and actuators used to control the motions of the links and rovers.

A novel two-stage optimal control technique is derived in this section and the control system block diagram is provided in Figure 2a.

The first stage of the diagram generates the reference trajectories for the end-effectors corresponding to a given payload trajectory. The Impedance control equations representing this first stage are provided in (18). These equations are developed in [2].

$$\ddot{X}_i = \underline{M}_i^{-1} \underline{B} \{ \dot{X}_{di} - \dot{X}_i \} + \underline{M}_i^{-1} \underline{K} \{ X_{di} - X_i \}, \quad i = L, R$$

(18)

where,  $M_i, K, B_i$ , are 6x6 positive definite matrices and are chosen in accordance with the tracking performance requirements.  $\tilde{X}_i$ . (i varies between L and R for each arm) are the end-effector trajectories while  $\tilde{X}_{di}$ . correspond to the reference trajectories of the attachment points on the common load.

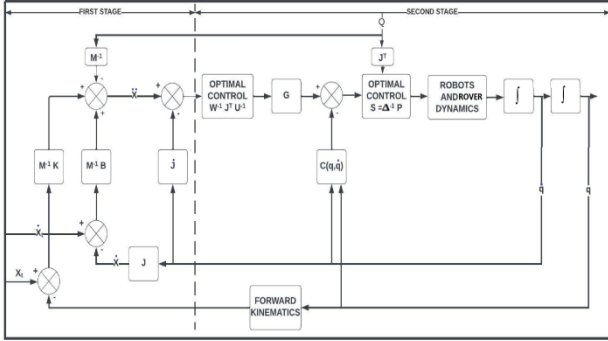


Figure 2: a. Optimal control system block diagram – Two Stage Control

The expressions for  $\tilde{X}_i$  and  $\dot{\tilde{X}}_i$  can be written as follows:

$$\tilde{X}_i = \begin{bmatrix} \tilde{\Omega}_k \\ \tilde{V}_k \end{bmatrix}^i \quad (19)$$

$$\dot{\tilde{X}}_i = \begin{bmatrix} \dot{\tilde{\Omega}}_k \\ \dot{\tilde{V}}_k \end{bmatrix}^i \quad (20)$$

where  $k$  point is the end-effector for the  $i^{\text{th}}$  arm. Performing the least-square minimization of joint rates, the inverse kinematics of the robotics system can be solved as the following:

$$\begin{bmatrix} \tilde{\Omega}_{ci} \\ \dot{\tilde{\theta}}_i \end{bmatrix} = \underline{W}_i^{-1} \left( \underline{J}_c^k \right)_i^T \underline{U}_i^{-1} \dot{\tilde{X}}_i \quad (21)$$

$$\begin{bmatrix} \dot{\tilde{\Omega}}_{ci} \\ \ddot{\tilde{\theta}}_i \end{bmatrix} = \underline{W}_i^{-1} \left( \underline{J}_c^k \right)_i^T \underline{U}_i^{-1} \left\{ \dot{\tilde{X}}_i - \left( \underline{J}_c^k \right)_i \begin{bmatrix} \tilde{\Omega}_{ci} \\ \dot{\tilde{\theta}}_i \end{bmatrix} \right\} \quad (22)$$

$$\underline{U}_i = \left( \underline{J}_c^k \right)_i \underline{W}_i^{-1} \left( \underline{J}_c^k \right)_i^T \quad (23)$$

where  $\underline{W}_i$  is a square positive definite weighting matrix with the dimensions of  $(n+3)$  by  $(n+3)$ .

The second stage in the block diagram is predicated on optimal control allocation (OCA). The mathematical model is provided below.

The performance index (a cost function)  $C$  is formulated to minimize the wheel moments  $\tilde{M}_L, \tilde{M}_R$  and the joint torques  $\tilde{\tau}_{\theta_L}, \tilde{\tau}_{\theta_R}$ , and the contact moments and forces  $\tilde{N}_i$  and  $\tilde{F}_i$  applied by the end-effectors on the common load

The performance index  $C$  can be expressed as:

$$C = \frac{1}{2} \tilde{S}^T \underline{H} \tilde{S} + \tilde{\lambda}^T \tilde{E} \quad (24)$$

$\underline{H}$  is a  $(2n+18, 2n+18)$  positive definite weighting matrix,  $\tilde{\lambda}$  is the  $((2n+12), 1)$  Lagrangian multiplier and  $\tilde{E}$  vector includes the equations of constraints and can be calculated as shown in (26).

The  $\tilde{S}$  vector contains the contact forces / moments, the wheel moments as well as the joint torques for the two arms and rovers as described below:

$$\tilde{S}^T = [\tilde{Q}_L, \tilde{Q}_R, \tilde{M}_L, \tilde{M}_R, \tilde{\tau}_{\theta_L}, \tilde{\tau}_{\theta_R}]$$

$$\tilde{Q}_i = \begin{bmatrix} \tilde{N}_i \\ \tilde{F}_i \end{bmatrix} \quad (25)$$

The  $\tilde{E}$  vector is provided below:

$$\tilde{E} = \begin{bmatrix} m_t \ddot{x}_t - \tilde{F}_L - \tilde{F}_R \\ [L_t \tilde{\Omega}_t + \tilde{\Omega}_t \times L_t \tilde{\Omega}_t] - [\tilde{N}_L + \tilde{N}_R - \tilde{d}_L \times \tilde{F}_L - \tilde{d}_R \times \tilde{F}_R] \\ G_{W_L} \quad G_{W_{LR}} \quad G_{W_{\theta_L}} \quad G_{W_{\theta_R}} \\ G_{W_{LR}}^T \quad G_{W_R} \quad G_{W_{\theta_L}} \quad G_{W_{\theta_R}} \\ G_{W_{\theta_L}}^T \quad G_{W_{\theta_L}} \quad G_{\theta_L} \quad G_{\theta_{LR}} \\ G_{W_{\theta_R}}^T \quad G_{W_{\theta_R}} \quad G_{\theta_{LR}} \quad G_{\theta_R} \end{bmatrix} \begin{bmatrix} \ddot{\Phi}_L \\ \ddot{\Phi}_R \\ \ddot{\theta}_L \\ \ddot{\theta}_R \end{bmatrix} - \begin{bmatrix} \tilde{c}_L \\ \tilde{c}_R \\ \tilde{c}_{\theta_L} \\ \tilde{c}_{\theta_R} \end{bmatrix} - \begin{bmatrix} \tilde{M}_L \\ \tilde{M}_R \\ \tilde{\tau}_{\theta_L} \\ \tilde{\tau}_{\theta_R} \end{bmatrix} \quad (26)$$

where  $\ddot{x}_t, m_t$  are the translational acceleration and the mass of the common load, respectively and.  $\tilde{d}_i^T = (x_i, y_i, z_i)$  is the position vector measured from the  $i^{\text{th}}$  arm's contact point to the load's mass center, while  $\tilde{\Omega}_t$  and  $L_t$  are the angular rate and the inertia matrix of the common load around its center of mass.

Once can minimize the performance index  $C$  by taking the derivative of  $C$  with respect to  $\tilde{\lambda}_i$  and  $\tilde{S}$  to obtain the minimum norm of wheel moments, joint torques, as well as the contact moments /force exerted by the end-effectors on the common load.

$$\frac{\partial C}{\partial \tilde{S}} = \tilde{0} \quad (27)$$

and

$$\frac{\partial C}{\partial \tilde{\lambda}} = \tilde{0} \quad (28)$$



One can obtain the minimum norm of  $\tilde{\mathcal{S}}$  containing the wheel moments, joint torques, as well as the contact force and moments by making use of the equations (27) and (28).

$$\tilde{\mathcal{S}} = \underline{\Delta}^{-1} \tilde{\mathcal{P}} \tag{29}$$

where  $\underline{\Delta}$  is a  $((2n + 18), (2n + 18))$  square matrix and  $\underline{\Delta}$  and  $\tilde{\mathcal{P}}$  are presented as follows:

$$\underline{\Delta} = \begin{bmatrix} \underline{H}_{NL} & \underline{0} & -\underline{H}_{NR} & \underline{0} & -\underline{H}_{rL} (J_c^k)_L^T & \underline{W}_{rR} (J_c^k)_R^T \\ (\underline{D}_L - \underline{D}_R) \underline{H}_{NL} & \underline{H}_{FL} & \underline{0} & -\underline{H}_{FR} & (\underline{D}_R - \underline{D}_L - \underline{1}) (J_c^k)_L^T & \underline{W}_{rR} (J_c^k)_R^T \\ \underline{0} & \underline{1} & \underline{0} & \underline{1} & \underline{0} & \underline{0} \\ \underline{1} & -\underline{D}_L & \underline{1} & \underline{D}_R & \underline{0} & \underline{0} \\ (J_c^k)_{L1}^T & (J_c^k)_{L2}^T & \underline{0} & \underline{0} & \underline{1} & \underline{0} \\ \underline{0} & \underline{0} & (J_c^k)_{R1}^T & (J_c^k)_{R2}^T & \underline{0} & \underline{1} \end{bmatrix} \tag{30}$$

$$\tilde{\mathcal{P}} = \begin{bmatrix} \underline{0} \\ \underline{0} \\ \underline{0} \\ \underline{0} \\ m_t \ddot{x}_t \\ \underline{L}_t \dot{\tilde{\Omega}}_t + \tilde{\Omega}_t \times \underline{L}_t \tilde{\Omega}_t \\ \begin{bmatrix} G_{WL} & G_{WLR} & G_{W\theta L} & G_{W\theta R} \\ G_{WL}^T & G_{WR} & G_{W\theta L} & G_{W\theta R} \\ G_{W\theta L}^T & G_{W\theta L} & G_{\theta L} & G_{\theta LR} \\ G_{W\theta R}^T & G_{W\theta R} & G_{\theta LR}^T & G_{\theta R} \end{bmatrix} \begin{bmatrix} \ddot{\Phi}_L \\ \ddot{\Phi}_R \\ \ddot{\Theta}_L \\ \ddot{\Theta}_R \end{bmatrix} \end{bmatrix} + \begin{bmatrix} \tilde{c}_L \\ \tilde{c}_R \\ \tilde{c}_{\theta L} \\ \tilde{c}_{\theta R} \end{bmatrix} \tag{31}$$

$$\underline{D}_i = \begin{bmatrix} 0 & -z_i & y_i \\ z_i & 0 & -x_i \\ -y_i & x_i & 0 \end{bmatrix} \tag{32}$$

### 2.5. Non-linear Model Predictive Control (NMPC) Technique

The control block diagram of the NMPC is illustrated in Figure.2b. It replaces the second stage of the model in Figure.2a. The reference trajectory shown in this diagram is the output of the first stage, i.e., the impedance control trajectory generation. However, in this case, the future state estimates are also taken into account to estimate the future reference trajectory values.

A robust NMPC algorithm is implemented by optimizing a performance index of the system which considers the predictions of the output signal and the constraints on the states, outputs and inputs as illustrated in Figure 2b.

The main difference between the Optimal Control Allocation (OCA) and the Non-linear Model Predictive Control (NMPC) is that the latter utilizes a model to predict and control future behavior, while the former only takes into account the current and the past.

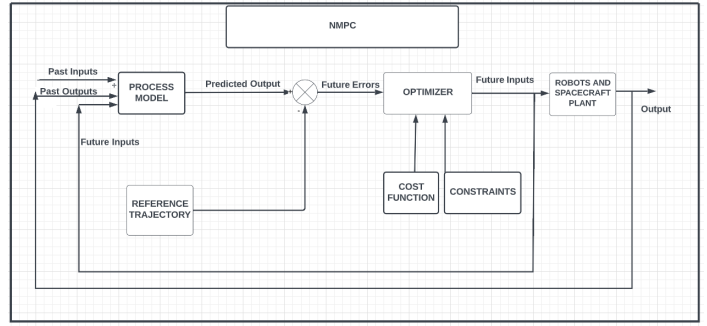


Figure 2: b Nonlinear Model Predictive Control (NMPC) block diagram

The optimization process carried out through the NMPC algorithm is performed at each control interval to predict the system's future behavior. It involves implementing various optimization problems related to the cost functions and constraints. The cost function is a type of scalar which needs to be minimized at intervals to assess the system's performance.

Besides the cost functions, the system also has to perform under various constraints to check its performance. These include the plant output and states. The modified states are adjusted depending on the constraints that are applied to the system.

The conventional MPC formulation for the multi-rover nonlinear system can be written as:

$$C = \int_0^{T_p} \left[ (\tilde{\mathcal{Y}}(t) - \tilde{\mathcal{Y}}_r(t))^T \underline{K} (\tilde{\mathcal{Y}}(t) - \tilde{\mathcal{Y}}_r(t)) + \tilde{\mathcal{S}}^T(t) \underline{H} \tilde{\mathcal{S}}(t) \right] dt$$

subject to:

$$\begin{aligned} \dot{\tilde{\mathcal{Y}}} &= \tilde{\mathcal{g}}(\tilde{\mathcal{Y}}) + \underline{L} \tilde{\mathcal{S}} \\ \tilde{\mathcal{z}} &= \tilde{\mathcal{g}}_z(\tilde{\mathcal{Y}}) + \underline{H} \tilde{\mathcal{S}} \\ \tilde{\mathcal{Y}}(0) &= \tilde{\mathcal{Y}}(t_0) \end{aligned} \tag{33}$$

where  $T_p$  is the prediction horizon;  $\underline{K}$  and  $\underline{H}$  are  $(2(2n+6) \times 2(2n+6))$  and  $((2n+18) \times (2n+18))$  positive definite square weighting matrices, respectively;  $\tilde{\mathcal{g}}(\tilde{\mathcal{Y}}), \tilde{\mathcal{g}}_z(\tilde{\mathcal{Y}}), \underline{L}, \underline{H}$  are part of the nonlinear system equations.

Also,  $\tilde{\mathcal{Y}}(t)^T = [q^T, \dot{q}^T]$ , is a  $(1 \times 2(6+2n))$  state vector,  $q^T$  vector is previously defined in Eq.(16), and  $\tilde{\mathcal{Y}}_r(t)$  is the reference / desired states.

The non-linear system can now be described with an exact quasi linear parameter varying realization:

$$\begin{aligned} \tilde{\mathcal{Y}}(k_t + 1) &= \underline{\hat{A}}(\hat{\mathcal{g}}(k_t)) \tilde{\mathcal{Y}}(k_t) + \underline{\hat{B}}(\hat{\mathcal{g}}(k_t)) \tilde{\mathcal{S}}(k_t) \\ \tilde{\mathcal{z}}(k_t) &= \underline{\hat{C}}(\hat{\mathcal{g}}(k_t)) \tilde{\mathcal{Y}}(k_t) + \underline{\hat{D}}(\hat{\mathcal{g}}(k_t)) \tilde{\mathcal{S}}(k_t) \\ \hat{\mathcal{g}}(k_t) &= \mathbf{f}_g(\tilde{\mathcal{Y}}(k_t)) \end{aligned} \tag{34}$$

where  $k_t$  is the sampling instant and  $\tilde{\mathcal{z}}(k_t)$  is a vector of the measured outputs at instant  $k_t$ .

The NMPC is employed at each sampling instant  $k_t$ . and the discrete states  $\tilde{\mathcal{Y}}(k_t)$  and control inputs  $\tilde{\mathcal{S}}(k_t)$  are obtained

minimizing the following performance index i.e., the Cost Function:

$$C = \frac{1}{2} \sum_{j=1}^{N_p} \left[ \begin{array}{c} (\tilde{\mathbf{y}}(k_t, +j) - \tilde{\mathbf{y}}_r(k_t, +j))^T \underline{K} (\tilde{\mathbf{y}}(k_t, +j) - \tilde{\mathbf{y}}_r(k_t, +j)) \\ + \tilde{\mathbf{S}}(k_t, +j - 1)^T \underline{H} \tilde{\mathbf{S}}(k_t, +j - 1) \end{array} \right]$$

subject to

$$\tilde{\mathbf{y}}(k_t + j + 1) = \underline{\mathbf{A}}(\hat{\mathbf{g}}(k_t + j))\tilde{\mathbf{y}}(k_t + j) + \underline{\mathbf{B}}(\hat{\mathbf{g}}(k_t + j)) \tilde{\mathbf{S}}(k_t + j)$$

$$\tilde{\mathbf{z}}(k_t + j) = \underline{\mathbf{C}}(\hat{\mathbf{g}}(k_t + j))\tilde{\mathbf{y}}(k_t + j) + \underline{\mathbf{D}}(\hat{\mathbf{g}}(k_t + j)) \tilde{\mathbf{S}}(k_t + j) \tag{35}$$

### 3. Computer Simulation Results and Discussion

The results of the computer simulations and their discussions are presented in this section. First, a prescribed trajectory for the common payload's center mass is generated. Then, the desired (reference) trajectories for the two end-effectors are obtained using a method known as the impedance control technique (shown as the first stage in the block diagram).

The goal of the simulation is to obtain the minimum joint and rover wheel torques and contact forces while simultaneously tracking the desired end-effector pose using the developed two different control algorithms (i) OCA and (ii) NMPC.

The parameters for the robotic arms and the rovers employed in the computer simulations are presented in Table 1. A mini version of the SSRMS is utilized.

Table 1: The System Parameters Utilized in the Computer Simulation

Description of Hardware Configuration Items	Dimensions (m)	Mass (kg)
Rovers- (#1 and #2)	(0.5x0.5x0.3)	40
Common Load	(0.4x1x0.4)	10
Link #1	(0.1x0.1x0.1)	1
Link #2	(0.1x0.1x0.1)	1
Link #3	(1x0.1x0.1)	3
Link #4	(1x0.1x0.1)	5
Link #5	(0.1x0.1x0.1)	3
Link #6	(0.1x0.1x0.1)	1
Link #7	(1x0.1x0.1)	3

The desired trajectories for the rotational and translational motions of the common load are presented with time in Figure 3.

The minimum norm of the contact moments /forces, the joint torques, as well as the control forces and moments on Rovers 1 and 2 are plotted in Figure 4a-m using OCA and NMPC algorithms. The non-optimum joint torques (in blue), the joint torques realized by application of OCA algorithm (in red) and by NMPC algorithm (in yellow) are plotted for comparison purposes. The comparison

of the plots illustrates that the NMPC is superior and then followed by OCA.

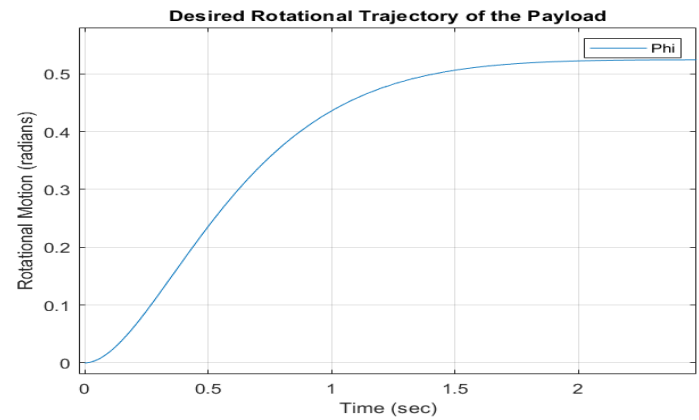
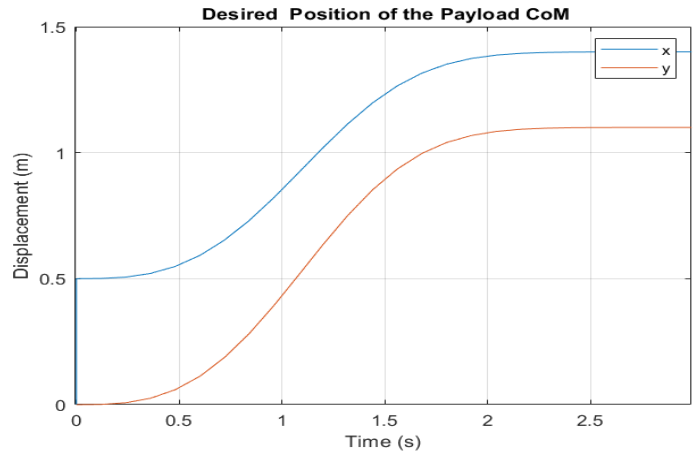


Figure 3: Variation of the reference trajectory for the common load

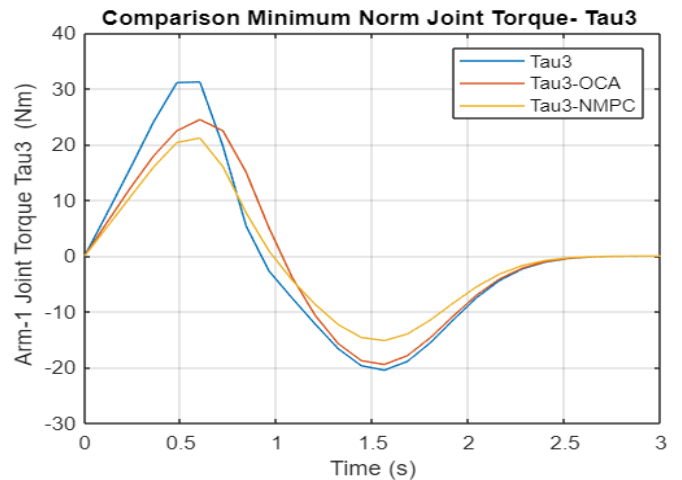


Figure 4: a Variation of the Joint 3 Torque obtained by Non-optimal, OCA, NMPC Algorithms (Arm 1)

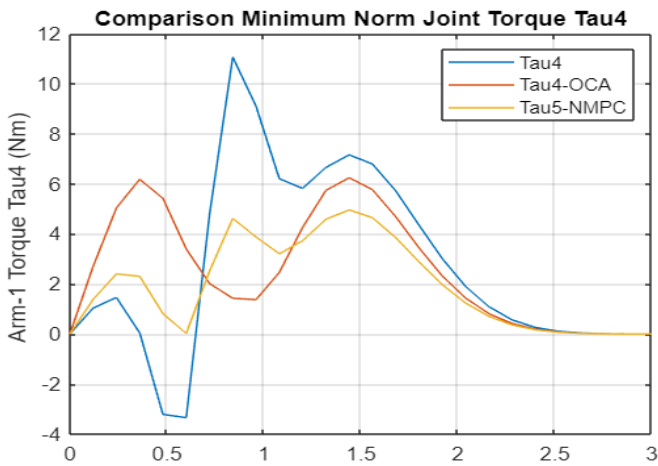


Figure 4: b- Variation of the Joint 4 Torque obtained by Non-optimal, OCA, NMPC Algorithms (Arm 1)

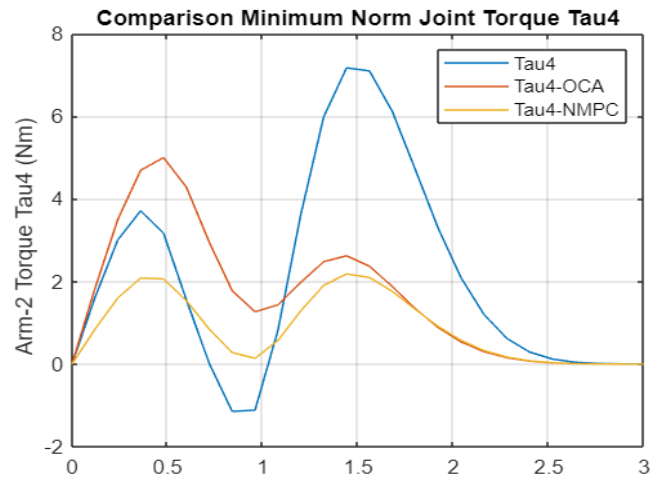


Figure 4: e- Variation of the Joint 4 Torque obtained by Non-optimal, OCA, NMPC Algorithms (Arm 2)

Comparison Minimum Norm Joint Torque Tau5 - In Orbit Plan

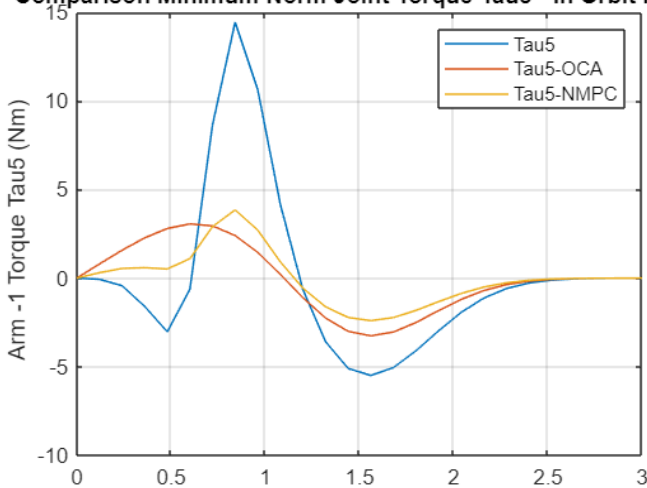


Figure 4: c- Variation of the Joint 5 Torque obtained by Non-optimal, OCA, NMPC Algorithms (Arm 1)

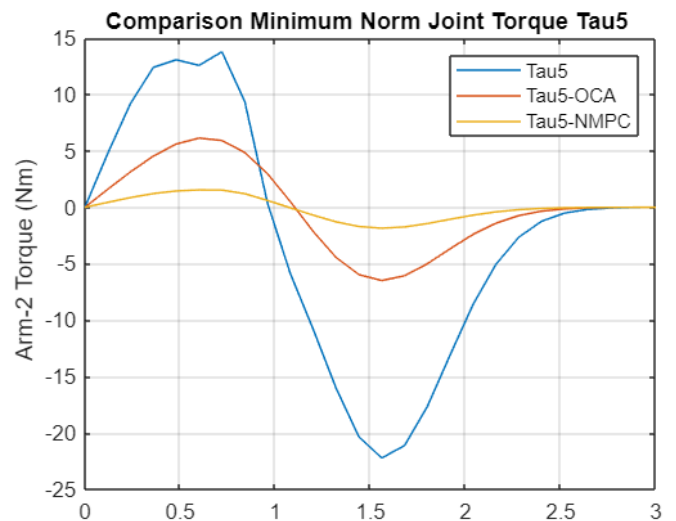


Figure 4: f- Variation of the Joint 5 Torque obtained by Non-optimal, OCA, NMPC Algorithms (Arm 2)

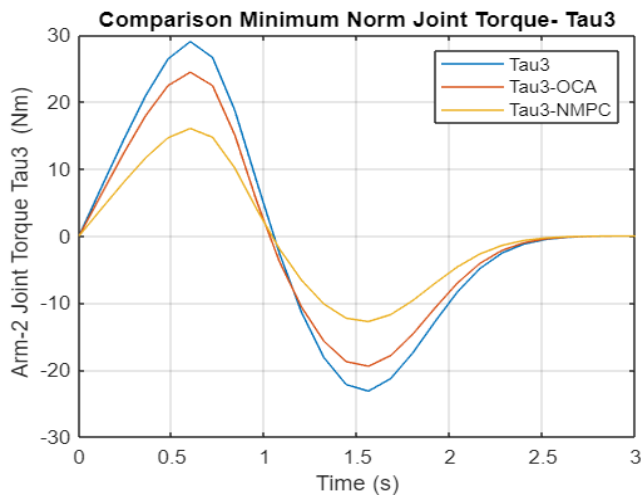


Figure 4: d- Variation of the Joint 3 Torque obtained by Non-optimal, OCA, NMPC Algorithms (Arm 2)

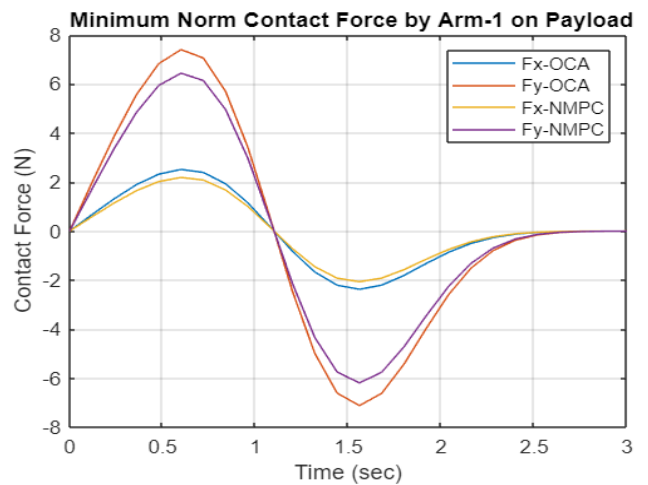


Figure 4: g- Variation of the Contact Forces on Payload by Non-optimal, OCA, NMPC Algorithms (Arm 1)

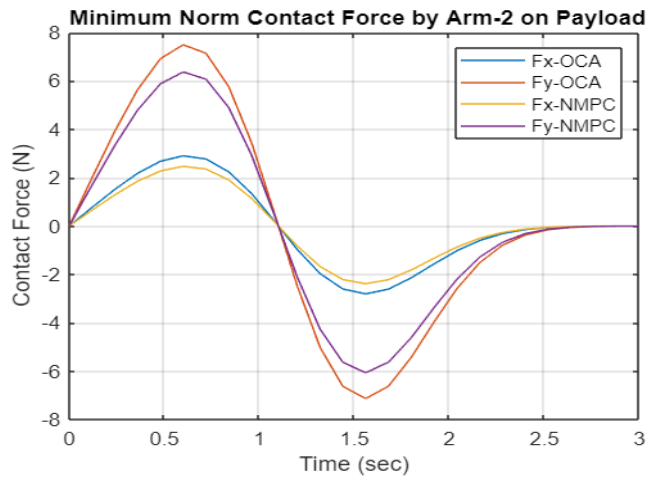


Figure 4: h Variation of the Contact Forces on Payload by Non-optimal, OCA, NMPC Algorithms (Arm 2)

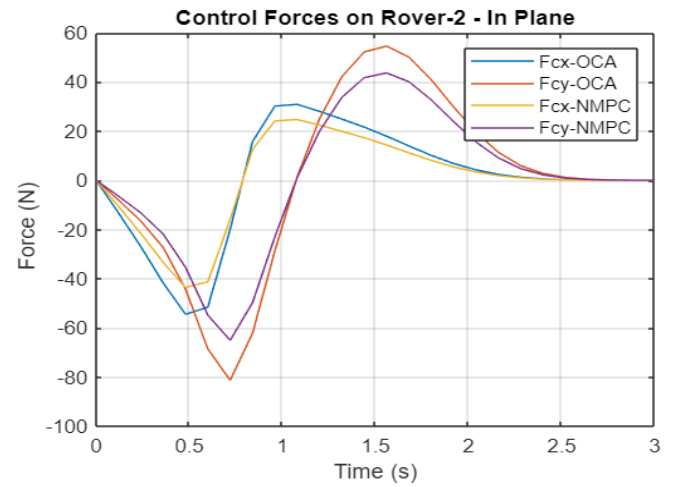


Figure 4: k Variation of the Control Forces on Rover 2 by Non-optimal, OCA, NMPC Algorithms (Rover 2)

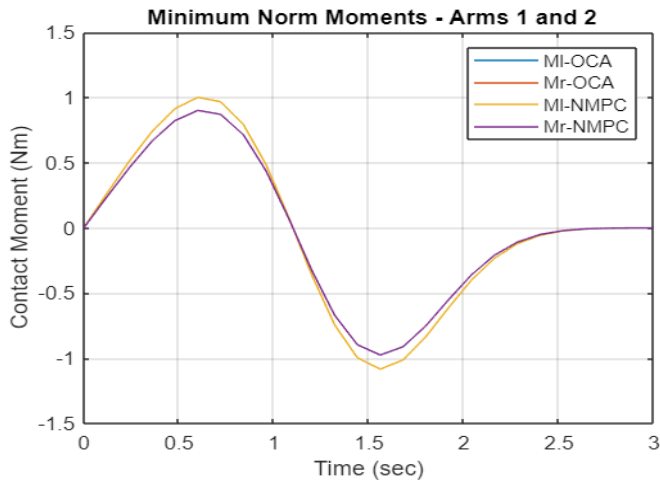


Figure 4: i Variation of the Contact Moments on Payload by Non-optimal, OCA, NMPC Algorithms (Arms 1 and 2)

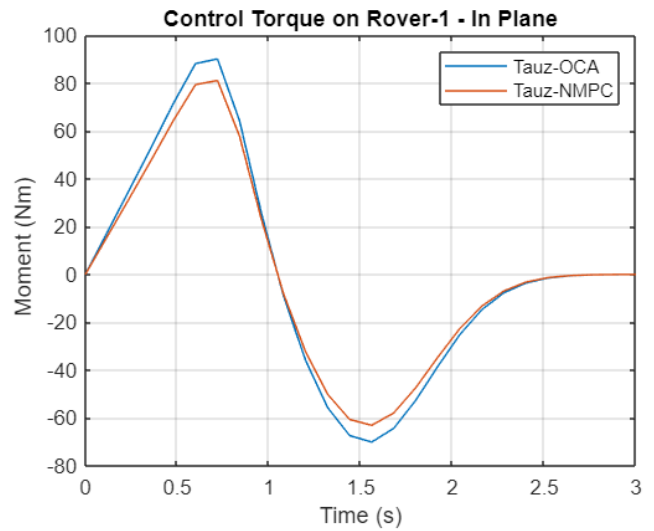


Figure 4: l Variation of the Control Moment on Rover 1 by Non-optimal, OCA, NMPC Algorithms (Rover 1)

Again, the NMPC is superior to OCA in obtaining minimum contact moments / forces applied to the common load while the two end-effectors are carrying a common load.

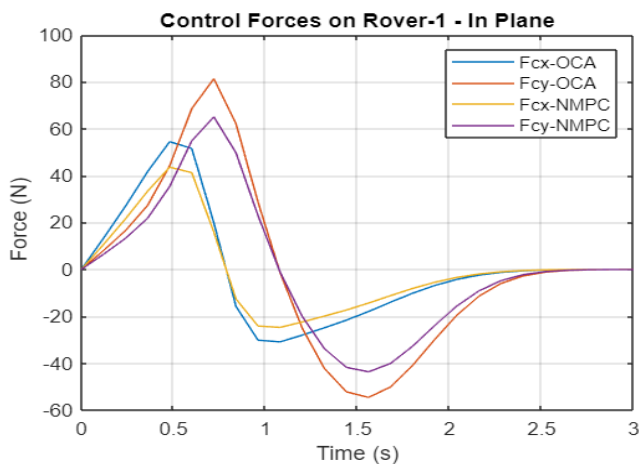


Figure 4: j Variation of the Control Forces on Rover 1 by Non-optimal, OCA, NMPC Algorithms (Rover 1)

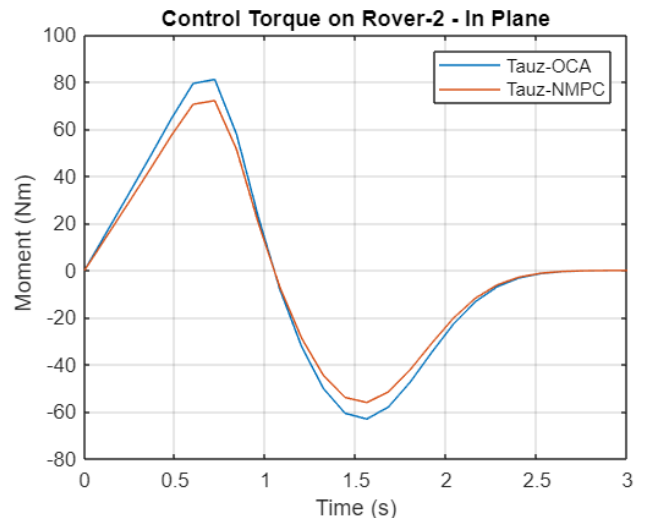


Figure 4: m Variation of the Control Moment on Rover 2 by Non-optimal, OCA, NMPC Algorithms (Rover 2)

A comparative analysis shows that again NMCP is superior to OCA in obtaining minimum norm of control moments and forces for Rovers 1 and 2.

The time variations of joint angular accelerations are shown in Figure 5.

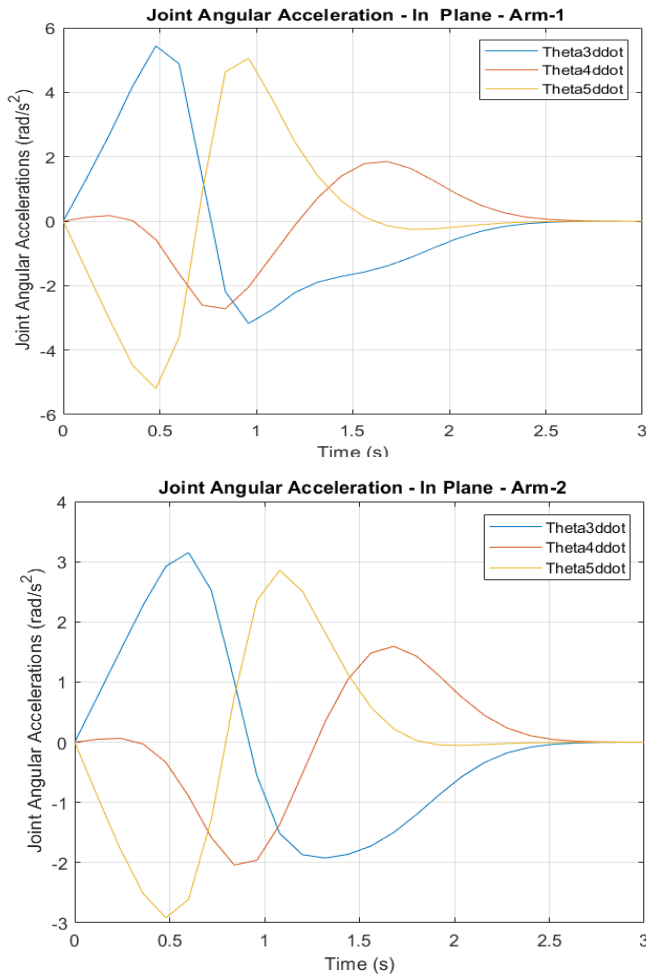


Figure 5: Variation of joint angular accelerations for the first and second arm

The joint accelerations are integrated to calculate rotational rates and angles using (13) and are presented in Figure 6.

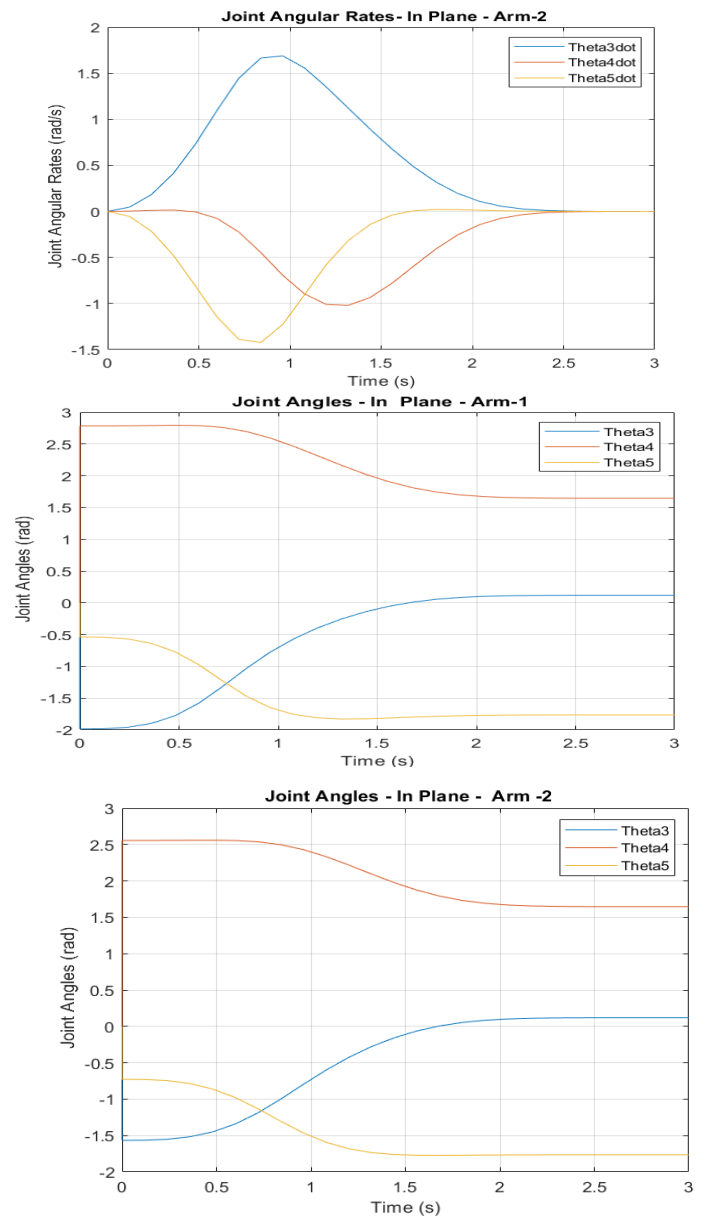
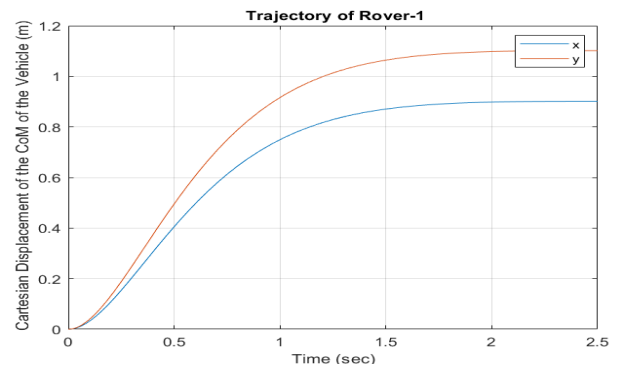
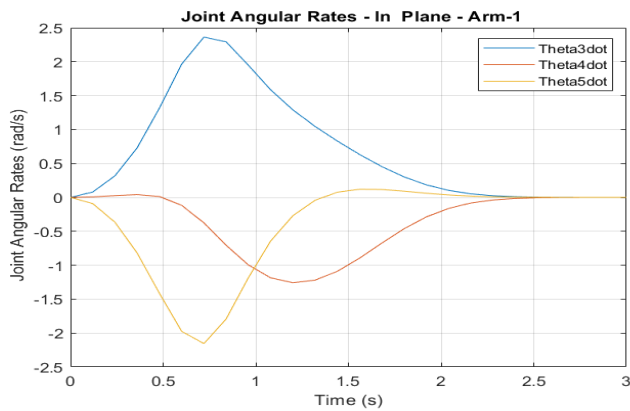


Figure 6: Variation of joint angular rates and angles for the first and second arm

The trajectories of the point C, the center of mass of the two rovers are determined by (22) and (8) and are shown in Figure 7.



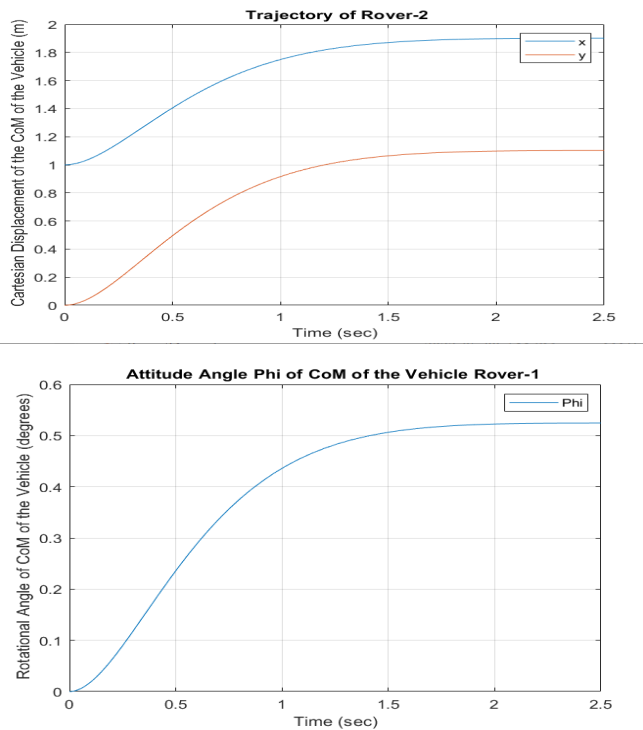


Figure 7: Variation of Rover 1 and 2 positions and orientations with time

The wheel angles of the two rovers are calculated utilizing (22) and are presented in Figure 8.

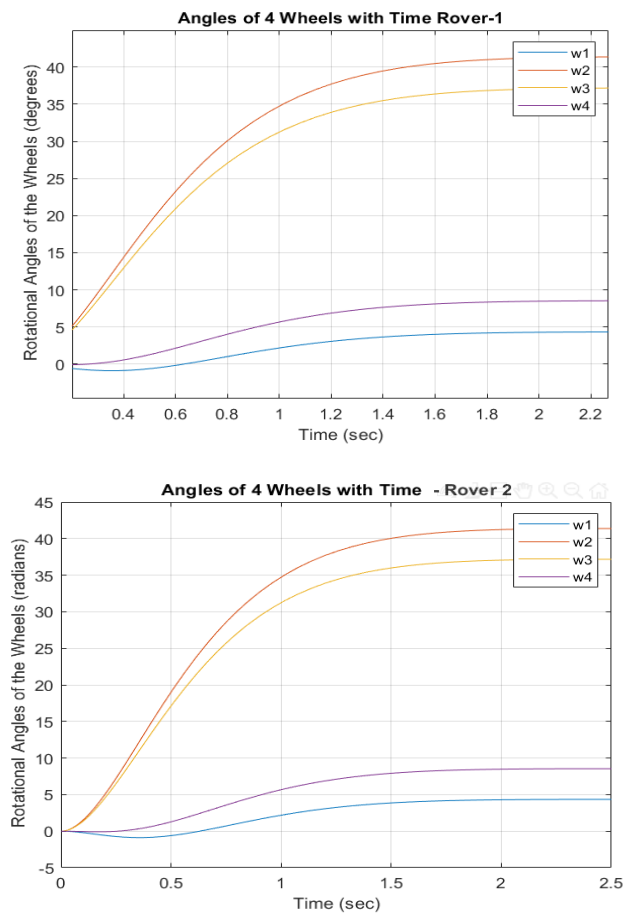


Figure 8 Variation of angles of rotations for rover wheels - rovers 1 and 2

## Conclusions and Future Work

The paper presented two novel control algorithms for motion and force control of a multi-rover robotics system when the two end-effectors carrying a common load. One algorithm is predicated on Optimal Control Allocation (OCA) and the other is a discretized (ii) Nonlinear Model Predictive Control (NMPC) algorithm.

The paper focused on developing robust and computationally efficient real-time control algorithms that can minimize the performance index consisting of the norm of the rovers control moments / forces, the joint torques, , as well as the contact moments / forces applied to the common load by two end-effectors.

The norm of wheel moments, joint torques, and the contact moments and forces were minimized to resolve the torque / moment saturation problem often seen while carrying a common load. The paper also presented a minimum norm solution for an underdetermined system subject to non-holonomic constraints. Moreover, the developed control algorithm also provided a real-time capability of trajectory for both the rovers and the arms while carrying a common load.

The system consisting of multi-rover with a dual arm was highly non-linear. The linear MPC technique on which most of the previous studies relied was not adequate. On the other hand, the computational complexity of a generic NMPC algorithm was very demanding. Therefore, in this paper, an elegant discretized technique with exact realization was implemented to take into account the full non-linear model and yet provide a simple real-time solution satisfying a minimum performance index subject to constraints.

The results of the computer simulations illustrated that the two algorithms OCA and NMPC worked efficiently. They were able to realize the minimum contact forces and moments and rover wheel moments and forces, joint torques, while manipulating a common load and tracking a reference load trajectory. In addition, the minimal norm solution also satisfied the non-holonomic constraints.

The results revealed that the optimization scheme used by the NMPC algorithm was the most effective when it came to achieving the lowest joint torques and forces. It was then followed by the OCA algorithm and the conventional least square method, respectively.

The authors are currently working on a research project to build a testbed to experimentally validate the computer simulation results. The comparisons of experimental and simulation results will be part of the future research work. Furthermore, the authors assumed no slippage occurred. However, the maximum driving force of each wheel is limited by the dynamic friction coefficient and the magnitude of the normal force acting on it. If this is exceeded, this assumption will no longer be valid. The normal forces will be incorporated in the dynamics model for the future work.

## Conflict of Interest

The authors declare no conflict of interest.

## References

- [1] S. Kalaycioglu, A. de Ruiter, "Coordinated Motion and Force Control of Multi-Rover Robotics System with Mecanum Wheels," in 2022 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), IEEE: 1–9, 2022, doi:10.1109/IEMTRONICS55184.2022.9795804.
- [2] D.S. Neculescu, B. Kim, S. Kalaycioglu, FREE MOTION, COLLISION AVOIDANCE AND CONTACT MOTION CONTROL FOR MOBILE ROBOTS, Elsevier: 223–228, 1993, doi:10.1016/B978-0-08-041897-1.50042-0.
- [3] N. Neculescu, B. Kim, S. Kalaycioglu, "Contact motion control for mobile robots," in 7th IFAC Symposium on Information Control Problems, IFAC, Toronto, 1992.
- [4] R. Fierro, F.L. Lewis, "Control of a nonholonomic mobile robot: backstepping kinematics into dynamics," in Proceedings of 1995 34th IEEE Conference on Decision and Control, IEEE: 3805–3810, doi:10.1109/CDC.1995.479190.
- [5] Yu Tian, N. Sidek, N. Sarkar, "Modeling and control of a nonholonomic Wheeled Mobile Robot with wheel slip dynamics," in 2009 IEEE Symposium on Computational Intelligence in Control and Automation, IEEE: 7–14, 2009, doi:10.1109/CICA.2009.4982776.
- [6] Y.H. Amengonu, Y.P. Kakad, "Dynamics and control for Constrained Multibody Systems modeled with Maggi's equation: Application to Differential Mobile Robots PartII," IOP Conference Series: Materials Science and Engineering, **65**, 012018, 2014, doi:10.1088/1757-899X/65/1/012018.
- [7] G. Campion, B. d'Andrea-Novell, G. Bastin, Controllability and state feedback stabilizability of non holonomic mechanical systems, Springer-Verlag, Berlin/Heidelberg: 106–124, doi:10.1007/BFb0039268.
- [8] A.M. Bloch, N.H. McClamroch, "Control of mechanical systems with classical nonholonomic constraints," in Proceedings of the 28th IEEE Conference on Decision and Control, IEEE: 201–205, doi:10.1109/CDC.1989.70103.
- [9] S. Kalaycioglu, "Control of multiple robot manipulators with optimal force distribution," in IEEE Canadian Conference on Electrical and Computer Engineering, 1991.
- [10] M. Vukob, S. Gros, G. Horn, G. Frison, K. Geebelen, J.B. Jørgensen, J. Swevers, M. Diehl, "Real-time nonlinear MPC and MHE for a large-scale mechatronic application," Control Engineering Practice, **45**, 64–78, 2015, doi:10.1016/j.conengprac.2015.08.012.
- [11] J.B. Rawlings, "Tutorial overview of model predictive control," IEEE Control Systems, **20**(3), 38–52, 2000, doi:10.1109/37.845037.
- [12] Y. Shi, K. Zhang, "Advanced model predictive control framework for autonomous intelligent mechatronic systems: A tutorial overview and perspectives," Annual Reviews in Control, **52**, 170–196, 2021, doi:10.1016/j.arcontrol.2021.10.008.
- [13] P.D. Christofides, R. Scattolini, D. Muñoz de la Peña, J. Liu, "Distributed model predictive control: A tutorial review and future research directions," Computers & Chemical Engineering, **51**, 21–41, 2013, doi:10.1016/j.compchemeng.2012.05.011.
- [14] M. Ellis, H. Durand, P.D. Christofides, "A tutorial review of economic model predictive control methods," Journal of Process Control, **24**(8), 1156–1178, 2014, doi:10.1016/j.procont.2014.03.010.
- [15] F. Michael, Implementation of Linear Model Predictive Control –Tutorial, 2021.
- [16] S. Yu, M. Reble, H. Chen, F. Allgöwer, "Inherent Robustness Properties of Quasi-infinite Horizon MPC," IFAC Proceedings Volumes, **44**(1), 179–184, 2011, doi:10.3182/20110828-6-IT-1002.01969.
- [17] H. Wei, C. Shen, Y. Shi, "Distributed Lyapunov-Based Model Predictive Formation Tracking Control for Autonomous Underwater Vehicles Subject to Disturbances," IEEE Transactions on Systems, Man, and Cybernetics: Systems, **51**(8), 5198–5208, 2021, doi:10.1109/TSMC.2019.2946127.
- [18] H. Wei, Q. Sun, J. Chen, Y. Shi, "Robust distributed model predictive platooning control for heterogeneous autonomous surface vehicles," Control Engineering Practice, **107**, 104655, 2021, doi:10.1016/j.conengprac.2020.104655.
- [19] K. Zhang, Q. Sun, Y. Shi, "Trajectory Tracking Control of Autonomous Ground Vehicles Using Adaptive Learning MPC," IEEE Transactions on Neural Networks and Adaptive Systems, **32**(12), 5554–5564, 2021, doi:10.1109/TNNLS.2020.3048305.
- [20] Y. Zou, X. Su, S. Li, Y. Niu, D. Li, "Event-triggered distributed predictive control for asynchronous coordination of multi-agent systems," Automatica, **99**, 92–98, 2019, doi:10.1016/j.automatica.2018.10.019.
- [21] K. Zhang, Y. Shi, "Adaptive model predictive control for a class of constrained linear systems with parametric uncertainties," Automatica, **117**, 108974, 2020, doi:10.1016/j.automatica.2020.108974.
- [22] J.S. Ladoiye, D.S. Neculescu, J. Sasiadek, "Force Control of Surgical Robot with Time Delay using Model Predictive Control," in Proceedings of the 15th International Conference on Informatics in Control, Automation and Robotics, SCITEPRESS - Science and Technology Publications: 202–210, 2018, doi:10.5220/0006908602020210.
- [23] R.A. Gangapersaud, G. Liu, A.H.J. de Ruiter, "Detumbling of a non-cooperative target with unknown inertial parameters using a space robot," Advances in Space Research, **63**(12), 3900–3915, 2019, doi:10.1016/j.asr.2019.03.002.
- [24] T. Englert, A. Völz, F. Mesmer, S. Rhein, K. Graichen, "A software framework for embedded nonlinear model predictive control using a gradient-based augmented Lagrangian approach (GRAMPC)," Optimization and Engineering, **20**(3), 769–809, 2019, doi:10.1007/s11081-018-9417-2.
- [25] K. Rathai, Synthesis and Real-time Implementation of Parameterized NMPC Schemes for Automotive Semi-active Suspension Systems, PhD Thesis, Communauté Universit'e Grenoble Alpes, Grenoble, 2020.
- [26] R. Quirynen, M. Vukob, M. Zanon, M. Diehl, "Autogenerating microsecond solvers for nonlinear MPC: A tutorial using ACADO integrators," Optimal Control Applications and Methods, **36**(5), 685–704, 2015, doi:10.1002/oca.2152.
- [27] F. Aghili, "Optimal control of a space manipulator for detumbling of a target satellite," in IEEE Int. Conf. Robot. Automatica, IEEE, 2009.
- [28] T. Rybus, J. Seweryn, J. Sasiadek, "Application of predictive control for manipulator mounted on a satellite," Archives of Control Sciences, **28**(1), 105–118, 2018.
- [29] M. Wang, J. Luo, U. Walter, "A non-linear model predictive controller with obstacle avoidance for a space robot," Advances in Space Research, **57**(8), 1737–1746, 2016, doi:10.1016/j.asr.2015.06.012.
- [30] M. Morato, J. Normey-Rico, O. Sename, "Model Predictive Control Design for Linear Parameter Varying Systems: A Survey," in Annual Reviews in Control, 64–80, 2020.
- [31] E. Psomiadis, E. Papadopoulos, "Model-Based/Model Predictive Control Design for Free Floating Space Manipulator Systems," in 2022 30th Mediterranean Conference on Control and Automation (MED), IEEE: 847–852, 2022, doi:10.1109/MED54222.2022.9837196.
- [32] M. Wada, S. Mori, "Holonomic and omnidirectional vehicle with conventional tires," in Proceedings of IEEE International Conference on Robotics and Automation, IEEE: 3671–3676, doi:10.1109/ROBOT.1996.509272.
- [33] J. Ostrowski, J. Burdick, "The Geometric Mechanics of Undulatory Robotic Locomotion," The International Journal of Robotics Research, **17**(7), 683–701, 1998, doi:10.1177/027836499801700701.
- [34] C. Stöger, A. Müller, H. Gattringer, Parameter Identification and Model-Based Control of Redundantly Actuated, Non-holonomic, Omnidirectional Vehicles, 207–229, 2018, doi:10.1007/978-3-319-55011-4\_11.
- [35] P.F. Muir, C.P. Neuman, "Kinematic modeling of wheeled mobile robots," Journal of Robotic Systems, **4**(2), 281–340, 1987, doi:10.1002/rob.4620040209.
- [36] F.G. Pin, S.M. Killough, "A new family of omnidirectional and holonomic wheeled platforms for mobile robots," IEEE Transactions on Robotics and Automation, **10**(4), 480–489, 1994, doi:10.1109/70.313098.
- [37] G. Campion, G. Bastin, B. D'Andrea-Novell, "Structural properties and classification of kinematic and dynamic models of wheeled mobile robots," in [1993] Proceedings IEEE International Conference on Robotics and Automation, IEEE Comput. Soc. Press: 462–469, doi:10.1109/ROBOT.1993.292023.
- [38] G. Wampfler, M. Salecker, J. Wittenburg, "Kinematics, Dynamics, and Control of Omnidirectional Vehicles with Mecanum Wheels," Mechanics of Structures and Machines, **17**(2), 165–177, 1989, doi:10.1080/15397738909412814.
- [39] A. Gfrerrer, "Geometry and kinematics of the Mecanum wheel," Computer Aided Geometric Design, **25**(9), 784–791, 2008, doi:10.1016/j.cagd.2008.07.008.
- [40] L.-C. Lin, H.-Y. Shih, "Modeling and Adaptive Control of an Omni-Mecanum-Wheeled Robot," Intelligent Control and Automation, **04**(02), 166–179, 2013, doi:10.4236/ica.2013.42021.

- [41] A. Shimada, S. Yajima, P. Viboonchaicheep, K. Samura, "Mecanum-wheel vehicle systems based on position corrective control," in 31st Annual Conference of IEEE Industrial Electronics Society, 2005. IECON 2005., IEEE: 6 pp., 2005, doi:10.1109/IECON.2005.1569224.
- [42] Y. Wang, D. Chang, "Motion performance analysis and layout selection for motion system with four Mecanum wheels," *Journal of Mechanical Engineering*, **45**(5), 307–316, 2009.
- [43] M.O. Tatar, C. Popovici, D. Mandru, I. Ardelean, A. Plesa, "Design and development of an autonomous omni-directional mobile robot with Mecanum wheels," in 2014 IEEE International Conference on Automation, Quality and Testing, Robotics, IEEE: 1–6, 2014, doi:10.1109/AQTR.2014.6857869.



# An Efficient Way of Hybridizing Edge Detectors Depending on Embedding Demand

Habiba Sultana\*, A. H. M. Kamal

Computer Science and Engineering, Jatiya Kabi Kazi Nazrul Islam University, Mymensingh, 2220, Bangladesh

## ARTICLE INFO

Article history:

Received: 29 August, 2022

Accepted: 01 January, 2023

Online: 24 January, 2023

Keywords:

Steganography

Edge detection

PSNR

## ABSTRACT

Edge detection-based image steganography schemes usually embed data in edge pixels only. However, some schemes embed data in non-edge pixels as well. In that case, the schemes embed more bits in the edges than in the smoothed areas. In all cases, the schemes perform large changes in a tiny area of the image during small data embedding. Detecting such local modifications is comparatively easier for a steganalyzer. As a result, it is preferable to distribute bits evenly across the image. Again, the schemes struggle to hide large messages in a cover image due to the algorithmic approach of hiding a fixed number of bits per pixel. In this research, we have overcome that problem by employing multiple edge detectors in generating a resultant edge image. Depending on the embedding needs, we have checked whether a single edge detector is sufficient to help in conceiving all bits or not. If it is not possible for a single-edge detector, we have hybridized them. Hybridization of edge images is performed either by logical AND, OR or OR with dilation. When the message size is very small, we have generated the resultant edge image by doing a logical AND operation among the edge images. That strategy have reduced the number of edge pixels as well as helped in distributing the to-be-embedded bits over the image in a more evenly manner. Similarly, to meet a larger embedding demand, we have performed a logical OR operation among the same edge images to increase the number of edge pixels. Even, to meet more embedding demand, we have dilated the OR-resultant image. These processes were carried out dynamically in the research according to an embedding demand. The experimental results deduce that this scheme embeds 92.37%, 73.92%, 74.78%, and 9.60% more bits than four competing methods. Similarly, for small embedding demand, the proposed scheme demonstrates 37.45%, 46.87%, 44.21%, and 55.56% higher PSNR values than the others. Moreover, statistical analyses state that this scheme demonstrates stronger security against attacks.

## 1 Introduction

In steganography, an embedding method implants a secret in a cover media such as a text file, digital image, audio, video, IP protocol, bio-signals, DNA sequence, etc [1]. By implanting secrets in a media, these methods modify the contents of that cover media. That modified media is then known as stego media. As a cover media, digital images are widely used in steganography because of their higher degree of redundant information [2]. The performance of image steganography is mainly measured by a set of features like payload, imperceptibility, and security of stego image [3]. Steganography methods work in either spatial domain, transform coefficients or created residues [1]. In the spatial domain, the confidential information is concealed in either pixel values or to their processed values [4]–[6]. There a very commonly used method is least signifi-

cant bit (LSB) substitution. In the transformed domain, the schemes first transform image contents, e.g., by wavelet transform, Fourier transform, etc., and then implant secrets in these coefficients [7]. In residue-based methods, the schemes implant secrets, generally, in pixel value differences (PVD) and prediction errors [8].

In terms of blindness, these schemes are categorized as reversible [6], [9]–[21] and irreversible [4, 5], [22]–[33] groups. In reversible steganography, the receiver rebuilds the cover image from the stego image in addition to extracting the desired secret message. On the other hand, the irreversible schemes extract the secrets only. Irreversible schemes are easy to implement and provide higher embedding capacity. For this, our research focuses on irreversible techniques. We concentrate our research target on the spatial domain only.

Machine learning is frequently used in cancer and kidney stone

\*Corresponding Author: Habiba Sultana, & Email: [srity.cse@gmail.com](mailto:srity.cse@gmail.com)

detection, image retrieval, and brain stroke [34]–[39]. Many of such applications use edge detection algorithm for localizing and visualizing target area in image and data. Before applying machine learning, if one wishes to implant privacy-preserving and security-related data [40]–[42] in the detected edge information that could be a promising technique to be used in the tele-medicine applications. Therefore, it is interesting to associate an edge detection method to divide the image contents into the edge and non-edge areas and to hide the data there [5, 6, 10, 31, 43].

In [22], [24], [33], [6], [2], the author used Canny edge detector to identify edge and non-edge pixels. All of the schemes used the LSB substitution method to hide data bits. In [25], [28] the author worked with different authors and applied a Canny edge detector in both cases to detect edge pixels. In [25] and [28], authors implanted data bits using reduplicated exploiting on the modification direction (REMD) and hybrid Hamming codes, respectively. In [8], the author employed a Canny edge detector as well and applied exclusive OR operation as a part of their embedding process. In [23], the author did the same but partitioned the image first into blocks. In addition to data implantation tasks, these schemes tried to their own ways to maintain a better visual quality in their stego images. In [3] and [31], the author used hybrid edge detection techniques in their data hiding process. They measured pixel value differences (PVD) first to decide the number of implanted bits per pixel and then used the LSB substitution method for data hiding. In [30], the author associated the Canny edge detector and PVD in their data hiding technique as well.

In [1] and [27], the author employed fuzzy edge detector to detect edge pixels and then used LSB substitution approach to implant the secrets into these detected edge pixels. In [1], the author used a chaotic method as a pre-processing task to encrypt the secret message.

Some authors first used multiple edge detectors using diverse edge operators, e.g., Canny, Sobel, and Fuzzy operators, and then hybridized these edge images to increase the number of edge pixels [4] in the resultant edge image. They used the LSB substitution method in their data implantation phase. To minimize the distortion of [4], in [26], the author divided the image into blocks and thereafter, they applied a hybridization process to these detected edge images. In [31], the author used another hybrid edge detection technique. He, additionally, performed a morphological dilation operator in their data-hiding phase to increase embedding capacity. In [32], the author hybridized the edge images by logical AND operator to increase the stego image quality while implanting small sized messages. In [5], the author proposed another edge detection based steganography method. They tested their scheme for Canny, Sobel, and Fuzzy-based edge detectors. They used the renown LSB substitution method to implant data bits. In [29], Ghosal proposed a steganography scheme using the Kirsch edge detection method where they implanted the message bits into each triplet of pixels. Therefore, the embedding capacity is low.

In this study, we have proposed a new hybrid edge detection-based embedding process where it embeds more bits in edge pixels than non-edge ones. The proposed scheme employs multiple edge detection methods and finds the best detector for what the demanded message bits are just conceivable. Depending on embedding demand, it determines whether a single edge detector is capable to

help in hiding entire secrets. If the size of the secret message is too small or very high than the embedding capacity by using a single edge detector, it hybridizes the edge images in both cases. When the message length is too small, the proposed scheme hybridizes edge images by logical AND operator to reduce the number of edge pixels. The number of edge images is employed in the AND operation depending on the message length. The resultant edge image helps the embedding algorithm to distribute the secret bits in the cover image more evenly. Again, to implant a large-size message, it performs a logical OR operation among the edge images to increase the number of edge pixels in the resultant edge image. Even, if OR is unable to implant the whole secret message, the scheme employs a morphological dilation operation to further increase the number of edge pixels in the resultant edge image. The scheme does that hybridization and dilation operation dynamically realizing the length of the secret message and computing the embeddable bits through that resultant edge image. Experimental results show that our proposed scheme performs better than the other competing methods for all the performance measuring parameters.

Contributions of this research are listed below:

- This scheme dynamically chooses the best one of the four different embedding techniques depending on the message length.
- We have allowed the scheme to implant a different number of bits in edge and non-edge pixels according to embedding demand.
- Our proposed method increases the visual quality of stego image and embedding capacity as well. At the same time, it shows strong resistance against statistical attacks.

The rest of this paper is organized as follows: section 2 concisely presents the related works. The proposed method is described neatly in section 3. Section 4 demonstrates the simulated results of our scheme. The results of testing the robustness of the proposed scheme against attacks are devoted in section 5. Finally, section 6 concludes the article.

## 2 Related Works

### 2.1 A Brief on Edge Detectors

The sharp changes in the image brightness are called the edges or boundaries of the image. Edges may exist in horizontal, vertical, or diagonal directions. The method which is used to detect the edges of an image is called edge detection. A filter, known as kernel or operator is used to identify the edges in an image. Very commonly used edge detectors are canny, sobel, log, Prewitt, kirsch, laplacian, and fuzzy. Generally, edge detectors are used in pattern recognition, feature extractions, and image morphology. In the field of detection-based, edge detectors are used to improve the security of data hidden. These schemes first detect the edge pixels and non-edge pixels in a cover image and implant only edge pixels or both categories. We have studied a good number of articles on that state art. Among those, we found the works of [5], [31], [33] and read them very

carefully and attentively and built the foundation of our proposed work on them.

### 2.2 Single edge-based image steganography

In [5], the author proposed an edge-detection-based steganographic method. This scheme copied the cover image and cleared  $n$ -bits LSBs from it. Then it applies various popular edge detectors such as canny, Sobel, fuzzy, etc., and generates an edge image. This scheme classifies the contents of cover image pixels as edge and non-edge pixels based on that edge image. This scheme then implants  $x$  bits of secrets into edge pixels and  $y$  bits into non-edge pixels and generates a stego image. This is the embedding process. In the extraction phase, the receiver extracts the secret messages from the stego image using the reverse process of embedding.

### 2.3 Hybrid edge-based image steganography

In 2018, Rasol [33] proposed an image steganography using a hybrid edge detection technique. In this scheme, the authors apply canny and Sobel edge detection methods and generate edge area. They combined those edge areas using logical OR operation. On the other hand, they also add a special character at the end of the message and convert it into binary according to ASCII. Then they embed  $x$  bits into the edge area and  $y$  bits into the non-edge area using the LSB method and thus generate a stego image and send it to the receiver. The receiver performs the reverse of the embedding process and extracts the secret message.

### 2.4 Dilated hybrid edge-based image steganography

In [31], the author proposed a dilated hybrid edge detection-based image steganography scheme. This method has three phases such as preprocessing, embedding, and extraction. Like as [5], this scheme also copies the cover image and cleared  $n$ -bits LSBs from it. This scheme then applies  $m$ -number of edge detectors such as  $e_1, e_2, \dots, e_m$  and combine two edge detectors using logical OR operation. This scheme also used morphological operators such as dilation to increase the number of edge pixels. Based on the dilated hybrid edge image, all the contents of the cover image pixels are classified as edge pixels and non-edge pixels. The XOR operator is used in the embedding process to improve security. This scheme implants  $x$  bits information in edge pixels and  $y$  bits into non-edge pixels and generates a stego image. In the extraction phase, the receiver extracts the secret message using the reversible method of embedding.

### 2.5 Edge-based image steganography

In [32], the author proposed an image steganography method based on hybrid edge detection. This scheme all most similar to setiadi's method [31]. The difference between those schemes is, [32] is applicable for small message sizes with maintaining good visual quality and [31] has good embedding capacity with maintaining visual quality. The scheme [32] used logical AND operation instead of logical OR operation.

Table 1 gives a summary of this work.

Table 1: Summary of related works. Uses of multiple edge detectors and dilating the hybrid edge image are the key distinguishing features of this scheme.

Criteria	[32]	[5]	[33]	[31]
Cleared LSBs?	Yes	Yes	No	Yes
Hybridize edge images?	Yes	No	Yes	Yes
Dilate edge image?	Yes	No	No	Yes
Use an edge detector?	Yes	Yes	Yes	Yes
Encrypt message?	No	No	Yes	Yes
Embed as (x,y) bits?	Yes	Yes	Yes	Yes
Message type?	Text	Binary	Text	Text

## 3 Proposed method

The proposed work consists of three phases: pre-processing, data embedding, and data extracting. The description of these phases is given below:

### 3.1 Pre-processing phase

As our target is to implant the message in both edge and non-edge pixels, we perform single edge detectors or hybridize edge detectors based on message length  $m_L$  and maximum achievable payload. We take an instance of the cover image and cleared  $n$ - bits LSBs from it. Then we apply  $m$ - number of edge detectors. We select the best suitable edge detector in the following way:

$$rE = \begin{cases} E & \text{when } m_L > P_s \text{ and } m_L < P_h \\ AND(E_1, E_2 \dots E_m) & \text{when } m_L \leq P_s \\ OR(E_1, E_2 \dots E_m) & \text{when } m_L \geq P_h \\ dilation(E) & \text{otherwise} \end{cases} \quad (1)$$

We select single edge detector  $E$  when message length  $m_L$  is greater than the probable highest payload  $P_h$  and less than the probable smallest payload  $P_s$ . When message length  $m_L$  is less than  $P_s$  then we hybridize two or more edge images using AND operation. we also hybridize two or more edge images using OR operation when message length  $m_L$  is greater than the probable highest payload  $P_h$ . In another case, we used morphological operator dilation when the message length is large. Let the cover image is  $C$  and an instance of it by  $I$ . We first clear  $n$ -bits of LSBs from every pixel of  $I$  by equation(2).

$$I(i, j) = I(i, j) - f(I(i, j), 2^n); \quad (2)$$

where function  $f$  returns the remainder value when one divides  $I(i,j)$  by  $2^n$ .

We have applied  $m$ -number of edge detection operators, e.g., canny, sobel, fuzzy, Robert, Prewitt, log, etc on the cleared image  $I$  to detect edge pixels, separately. We have generated the edge image by equation (3).

$$eI(i) = \psi(I, \Omega); \quad (3)$$

where  $\psi$  is one of the  $m$  edge detection operators, i.e.,  $\Omega \in \{canny, sobel, log, fuzzy, Robert, Prewitt, etc.\}$  and  $1 \leq I \leq$

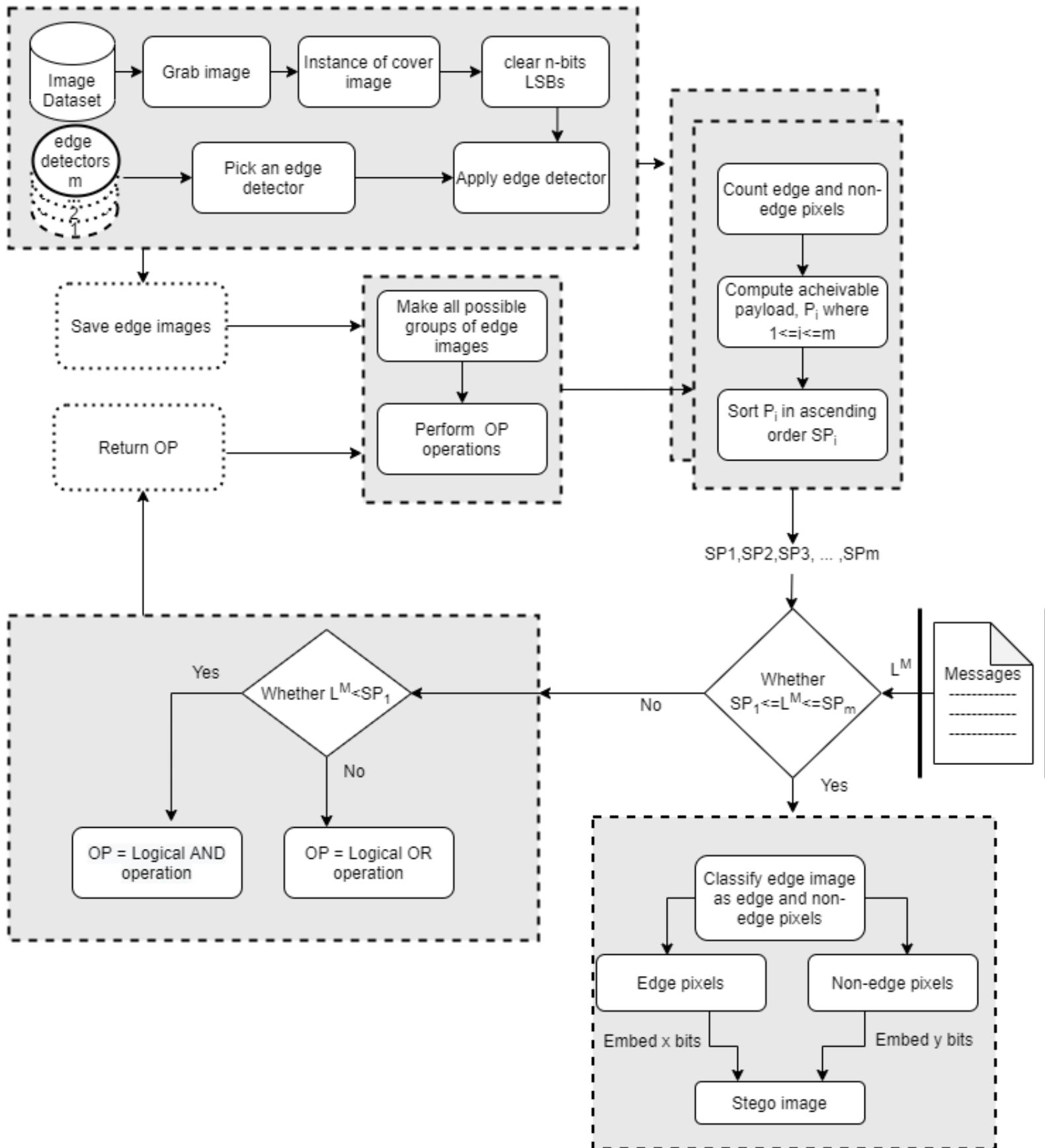


Figure 1: Preprocessing and embedding Phase.

$m$  and  $\psi$  returns the edge image  $eI$  from  $I$  for a specific edge detector  $\Omega$ . We also make all possible combinations of edge images if needed. We performed logical AND, OR, or dilation operations in each group. Each edge image is a binary image. For each pixel, the edge image holds a 0 or 1. A 1-in-edge image means the corresponding pixel of  $I$  is in the detected edge.

We compute the maximum achievable payload  $P_i$  by using equation (4). Payload is the total embedded bits.

$$P_i = ePN * x + neN * y; \tag{4}$$

Where  $ePN$  is the edge pixel and  $neN$  is the non-edge pixels and  $x$  and  $y$  are the number of embedded bits,  $1 \leq x \leq 5$  and  $1 \leq y \leq 4$ .

We then sort  $P_i$  in ascending order i.e.  $P_1 < P_2 < P_3 < \dots < P_i$ . We calculate the total embedded message length  $m_L$  by equation (5).

$$m_L = \text{length}(\text{message}); \tag{5}$$

Next, we check which one first meets the requirement of embedding payload  $m_L$ , say  $P_k$  using equation (1).  $P_k$  is the resultant edge image.

Those are the pre-processing stage. This is illustrated in Figure (1) up to (4) blocks.

### 3.2 Data embedding phase

Now it is the time to implant secrets in the image  $I$ . The pre-processing and embedding process is illustrated in Figure (1). The data implantation steps are as follows:

- An algorithm is developed to classify edge and non-edge pixels and their location in  $I$  based on resultant edge image  $P_k$  by equation (6).

$$[ePN, ePNP, neN, neNP] = F(I, P_k); \tag{6}$$

Where  $F$  returns edge pixels  $ePN$ , their positions  $ePNP$ , non-edge pixels  $neN$  and their positions  $neNP$  in  $I$ . The function  $F$  performs in following:

---

```
Function  $F(I, P_k)$ 
Compute the size of image  $I$ . Let it is  $(h, w)$ 
 $[ePN, ePNP] = G_e(I, P_k, h, w)$ 
 $CP_k = (P_k - 1) * (-1)$ 
 $[neN, neNP] = G_e(I, CP_k, h, w)$ 
return  $[ePN, ePNP, neN, neNP]$ 
```

---

Where  $G_e$  is defined below.

---

```
Function  $G_e(Q, R, h, w)$ 
 $k = 0$ 
for  $i = 1$  to  $h$ 
for  $j = 1$  to  $w$ 
if  $R(i, j) == 1$ 
 $R_1(k) = Q(i, j)$ 
 $R_2(k, 1) = i, R_2(k, 2) = j$ 
return  $R_1, R_2$ 
```

---

- Now, we implant  $x$  bits and  $y$  bits of secrets in each edge and non-edge pixel, respectively of the cover image by the

LSB substitution method. Let  $x$  bits of information be  $b_x$  and  $y$  bits of information are  $b_y$ . In that, the substitution task is performed by equation (7).

$$\begin{cases} S(s, t) = I(s, t) + \phi(b_x) \\ S(u, v) = I(u, v) + \phi(b_y) \end{cases} \tag{7}$$

where  $\phi(b_x)$  stands for decimal conversion of binary  $b_x$ .  $s = ePNP(i, 1)$ ,  $t = ePNP(i, 2)$ ,  $u = neNP(j, 1)$ ,  $v = neNP(j, 2)$  and  $1 \leq i \leq No\_O\_Edge\_Pixels$ ,  $1 \leq j \leq No\_Of\_nonEdge\_Pixels$ . Here  $b_x$  will be different for each of the  $s$  and  $t$ . The same is true for  $b_y$ . This means that each time a different  $b_x$  and  $b_y$  of the secret will be implanted. That stego image  $S$  is then sent to a receiver end. The receiver end next extracts the implanted secrets from  $S$ .

### 3.3 Data extraction phase

In the extraction phase, the receiver receives the stego image and stego key. The receiver gets the necessary information from the stego key such as the number of cleared bits  $n$ , the name of selected edge detectors, the number of bits embedded in edge and non-edge pixels, and message length. Like the sender, the receiver copies the stego image  $S$  to  $I$ . It then clears  $n$  bits of LSBs from  $I$  by equation (1). Let, that  $n$ -LSBs cleared image is also  $I$ . The scheme that applies  $m$ - number of edge detectors on  $I$  from stego key. We have then separated the edge and non-edge pixels and their corresponding positions in  $I$  by equation (6). Next from each of the edge and non-edge located pixels, i.e., from  $(s, t)$  and  $(u, v)$ , we have measured  $d_x$  and  $d_y$  using equations (8) and (9).

$$d_x = S(s, t) - I(s, t); \tag{8}$$

$$d_y = S(u, v) - I(u, v); \tag{9}$$

Here  $s, t, u, v, i, \text{ and } j$  are defined in the previous subsection. Next, we extracted the binaries of the secret by equation (10).

$$\begin{cases} b_x = \phi^{-1}(d_x) \\ b_y = \phi^{-1}(d_y) \end{cases} \tag{10}$$

Where  $\phi^{-1}$  means binary conversion of decimal value  $d_x$ .

## 4 Result analysis and discussion

In this section, we show the experimental results conducted to evaluate the performance of the proposed scheme with the works of Sultana [32], Bai [5], Rasol [33] and Setiadi[31]. We first selected ten frequently used images, an image dataset, and a message dataset. We set up our experiment and then analyzed the results.

### 4.1 Experimental Setup

We worked on MATLAB's edition R(2017a) on windows 7. The experiments were performed on a desktop that is specified by an Intel (R) Core (TM) i5-8500T CPU @ 2.10 GHz 2.11 GHz processor and RAM of 8.00 GB. In the proposed system we used two different types of input data one is the secret message, i.e. to be implanted data, and the other is the cover image. We first collected some sample messages from different sources, as shown in Table 2.

The sample message could be a text, binary, or any other format. We used our prepared function ConBin to convert the non-binary input data to binary. For example, text data is converted to binary from the ASCII values of the text contents. We work for different sizes of message lengths. As a cover media, We collected ten frequently used standard images as shown in Figure 2 to conduct all our primary experiments. We also used 499 images of the BOSS dataset. We converted the color of the images to grayscale and resized them to 512 x 512. As the contents of the dataset are images, we worked with intensity values of pixels. Thus the final inputted values to our embedding algorithm are binary for secret message and pixel values for cover media. We measured the performance of the schemes by several feature values, such as edge pixel generation capability, embedding capacity, peak signal-to-noise ratio (PSNR), structural similarity index matrix (SSIM), standard deviation, correlation coefficient, entropy, cosine similarity, and Pixel difference histogram, etc.

Table 2 is given as a message dataset of this work.

Table 2: List of message dataset with message length and type

DatasetName	Message Length	Message Type
SupervisorMessage	398833	text
M1	274661	binary
M2	442483	text
M3	693637	binary

A supervisor Message is a text-type message from my supervisor. The message is  
Though the life of a Ph.D. researcher is a matter of struggle, it is enjoyable as well. Bethe cause finding some novelty is always challenging and overcoming such challenge gives a researcher heavenly happiness.....

#### 4.2 Mathematical Representation of Feature values

Let the number of edge and non-edge pixels in a cover image are  $ePN$  and  $neN$ , respectively. Then the maximum achievable payload  $PL$  is defined by equation (11)

$$PL = ePN * x + neN * y; \tag{11}$$

Where  $x$  and  $y$  are the numbers of bits embedded in the edge and the non-edge pixels. We also measured the capacity. Capacity can be defined as the number of implanted bits per pixel. Embedding capacity  $EC$  is measured by equation (12)

$$EC = \frac{P}{h * w}; \tag{12}$$

Where,  $P$  is the total number of implanted bits in the cover image,  $h$  and  $w$  are the image height and width. Maintaining image quality is a challenging task and for this purpose, we used PSNR and SSIM which are commonly used image distortion measurement parameters. PSNR is measured by equation (13)

$$PSNR = 10 \log_{10} \frac{255^2}{MSE}; \tag{13}$$

where,

$$MSE = \frac{1}{h * w} \sum_{i=1}^w \sum_{j=1}^h (S_{i,j} - C_{i,j})^2; \tag{14}$$

here,  $S$  is the stego image and  $C$  is the original cover image. Next, SSIM is calculated by equation (15)

$$SSIM = \frac{(2\mu_c\mu_s + C_1)(2\sigma_{cs} + C_2)}{(\mu_c^2 + \mu_s^2 + C_1)(\sigma_c^2 + \sigma_s^2 + C_2)} \tag{15}$$

here,  $\mu_c$  and  $\sigma_c$  are the mean and variance of pixel values in the cover image. Likewise cover,  $\mu_s$ , and  $\sigma_s$  are the same for the stego image.  $C_1$  and  $C_2$  are two constants and we set  $C_1 = 0.0001$  and  $C_2 = 0.0009$  for experiment. There are many methods of analyzing the robustness against various attacks. Famous techniques are entropy measurement, standard deviation measurement, analyzing correlation among the pixels, checking the cosine similarity between the cover and stego image, and histogram of the Pixel difference between the stego and cover image. The entropy is measured by equation (16)

$$H = - \sum_k P_k \log_2(P_k); \tag{16}$$

Where,  $P_k$  is the probability associated with gray value  $k$  and  $1 \leq k \leq 255$ .

Standard deviation is defined by equation (17)

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}. \tag{17}$$

Where  $N$  is the number of data points,  $x_i$  each of the values of the data, and  $\bar{x}$  is the mean of  $x_i$ .

Population correlation is defined by equation (18)

$$P_{cs} = \frac{\sigma_{cs}}{\sigma_c \sigma_s}; \tag{18}$$

Where  $\sigma_c$  and  $\sigma_s$  are population standard deviations in cover  $C$  and stego  $S$ . Again,  $\sigma_{cs}$  is the co-variance between the cover and stego image. Equation (19) gives us the cosine similarity values

$$f_{\cos sim}(C, S) = \cos \theta = \frac{\sum_{i=1}^h \sum_{j=1}^w C(i, j)S(i, j)}{\sqrt{\sum_{i=1}^h \sum_{j=1}^w C(i, j)} \sqrt{\sum_{i=1}^h \sum_{j=1}^w S(i, j)}}; \tag{19}$$

Where  $C$  and  $S$  are cover and stego images.

#### 4.3 Experimental results and discussions

In the experiment, we first applied canny, sobel, log, Prewitt, and Roberts edge detectors in five LSBs cleared images to identify edge and non-edge pixels. Canny, sobel, log, Prewitt, and Roberts-based edge detector functions of MATLAB return an edge image for a given input image. The edge image is a binary image. The obtained edge images generated from ten input images, are shown in Table 3. In the previous section, according to the embedding rules we implant  $x$  bits of information in edge pixels and  $y$  bits of information in non-edge pixels and that represent as a tuple  $(x, y)$  where  $x > y$ . Table 4 summarises the number of edge pixels that were found in ten sample images by different methods. Table 4 provided statistics collected from 5-LSBs cleared images.

We calculated the maximum achievable payload of each edge image

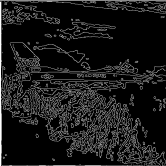
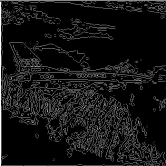












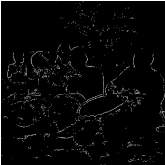
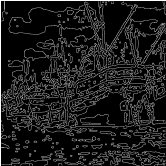






















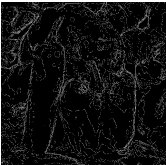


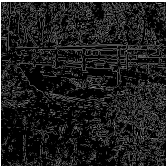





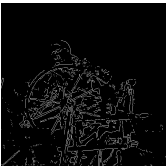

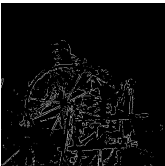


Figure 2: Sample cover images.



Figure 3: Stego images for the cover images of Figure 2.

Table 3: Edge images generated from ten cover images. The images were formed for Canny, Log, Prewitt, Sobel, and Roberts edge detectors from 5-LSBs cleared images (n=5).

ImageName	Methods				
	Canny	Log	Prewitt	Sobel	Roberts
F16.jpg					
babon.jpg					
basket.jpg					
boat.jpg					
brbra.jpg					
lena.jpg					
livingroom.jpg					
pepper.jpg					
walkbridge.jpg					
wheel.jpg					



which is shown in Figure 4. That figure states that canny is the highest and Prewitt is the lowest value of the maximum achievable payload. At the same time, for the experimental purpose, we take four sample messages and the length are 398833 bits, 274661 bits, 442483 bits, and 693637 bits. We demonstrate various experimental results for those messages.

We analyzed the performance in embedding capacity as well. Embedding capacity is graphically shown in Figure 5 for 499 images of the BOSS dataset.

We also analyzed the visual quality and structural originality of stego images. Visual quality is measured by PSNR values and is sketched in Figure 6 for a message length of 398833 bits. PSNR value in our scheme is higher than the others which are 37.45%, 46.87%, 44.21%, and 55.56%. It is clear from the diagram that the proposed scheme has a higher PSNR value than the competing schemes. The structural similarity index value, SSIM, for message length = 398833 bits is listed in Table 5. The table confirms that the SSIM values of all the schemes are both high and very close to each other.

Table 7 and Table 8 contain the values of PSNR and SSIM for message lengths 274661 bits, 442483 bits, and 693637 bits. We analyzed the time complexity of diverse schemes by measuring their required data embedment times for a message length of 398833 bits. The time complexity of the schemes is measured experimentally and tabulated in Table 6.

Table 4: A comparison of the number of edge pixels on various edge detectors for cleared images (n=5).

ImageName	Methods				
	Canny	Log	Prewitt	Sobel	Roberts
F16.jpg	24966	19532	8119	7786	5239
babon.jpg	38383	26700	2700	1966	26603
basket.jpg	31560	21967	8603	9145	5323
boat.jpg	26991	21406	4555	14018	23576
brbra.jpg	26941	19099	4536	4010	1595
lena.jpg	24884	19391	12270	12229	21061
livingroom.jpg	35543	25742	12253	11998	7049
pepper.jpg	25860	19595	4686	14211	21890
walkbridge.jpg	45563	28134	8405	9388	4685
wheel.jpg	27745	20467	8726	8870	8544

Table 5: SSIM values when message length = 398833 bits.

ImageName	SSIM values				
	Proposed	[32]	[5]	[33]	[31]
FF16.jpg	0.993	0.992	0.975	0.984	0.961
babon.jpg	0.972	0.971	0.924	0.946	0.911
basket.jpg	0.997	0.995	0.987	0.989	0.979
boat.jpg	0.993	0.962	0.964	0.980	0.948
brbra.jpg	0.994	0.995	0.966	0.986	0.945
lena.jpg	0.987	0.954	0.953	0.973	0.928
livingroom.jpg	0.997	0.994	0.986	0.988	0.978
pepper.jpg	0.990	0.950	0.954	0.979	0.933
walkbridge.jpg	0.998	0.996	0.986	0.989	0.984
wheel.jpg	0.996	0.994	0.989	0.990	0.982

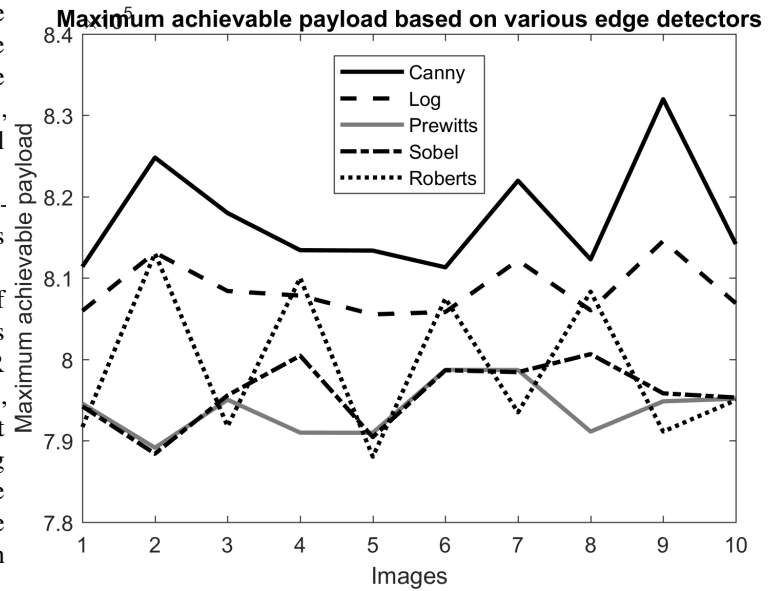


Figure 4: Maximum achievable payload based on diverse edge detectors where  $x = 4$  and  $y = 3$ . The figure states that canny has the highest achievable payload than log, Prewitt's, sobel, and Roberts.

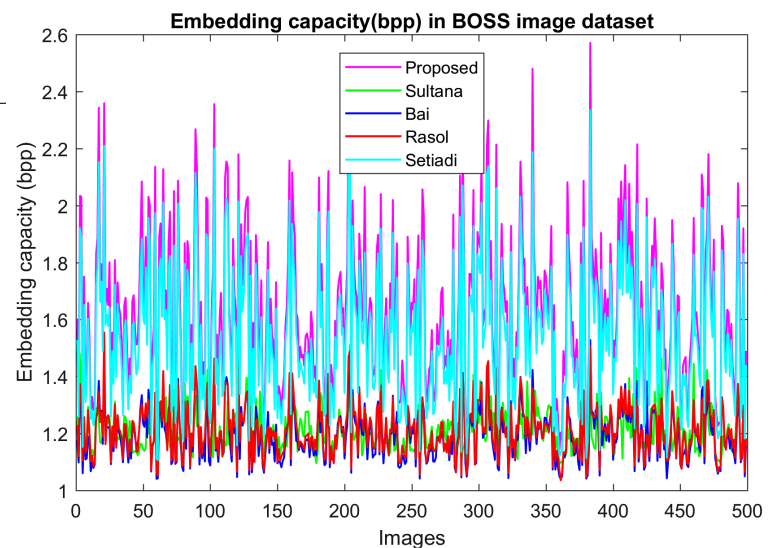


Figure 5: Performance comparison of the proposed scheme with Sultana [32], Bai [5], Rasol [33], and Setiadi [31] in terms of capacity in the BOSS image dataset.

Table 6: Elapsed time for message length = 398833 bits.

ImageName	Elapsed time				
	Proposed	[32]	[5]	[33]	[31]
F16.jpg	13.32	13.32	13.32	13.47	12.20
babon.jpg	13.45	13.23	13.66	13.75	9.89
basket.jpg	13.69	13.38	13.61	13.59	11.01
boat.jpg	13.28	13.41	13.50	13.68	11.18
brbra.jpg	13.18	13.00	13.49	13.32	11.37
lena.jpg	13.19	13.43	13.52	13.45	11.51
livingroom.jpg	13.48	13.39	13.58	13.72	9.72
pepper.jpg	14.38	13.39	13.47	13.47	11.26
walkbridge.jpg	13.60	13.27	13.81	13.85	9.44
wheel.jpg	13.49	13.42	13.43	13.46	11.64

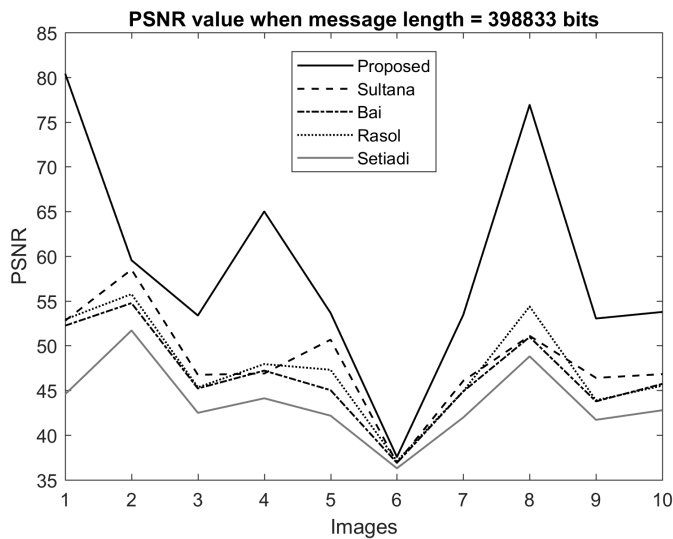


Figure 6: Performance comparison of the proposed scheme with Sultana [32], Bai [5], Rasol [33], and Setiadi [31] in terms of PSNR when message length = 398833 bits. The figure states that the proposed scheme dominates the other competing schemes.

Table 7: SSIM values of different schemes

Methods	ImageName	Message length (bits)		
		274661	442483	693637
Proposed method	F16.jpg	0.994	0.967	0.948
	boat.jpg	0.991	0.953	0.924
	pepper.jpg	0.988	0.941	0.905
[32]	F16.jpg	0.993	0.992	0.992
	boat.jpg	0.972	0.962	0.962
	pepper.jpg	0.962	0.950	0.950
[5]	F16.jpg	0.978	0.975	0.975
	boat.jpg	0.972	0.964	0.964
	pepper.jpg	0.963	0.954	0.954
[33]	F16.jpg	0.987	0.984	0.984
	boat.jpg	0.985	0.980	0.980
	pepper.jpg	0.982	0.979	0.979
[31]	F16.jpg	0.969	0.957	0.955
	boat.jpg	0.961	0.940	0.929
	pepper.jpg	0.953	0.924	0.913

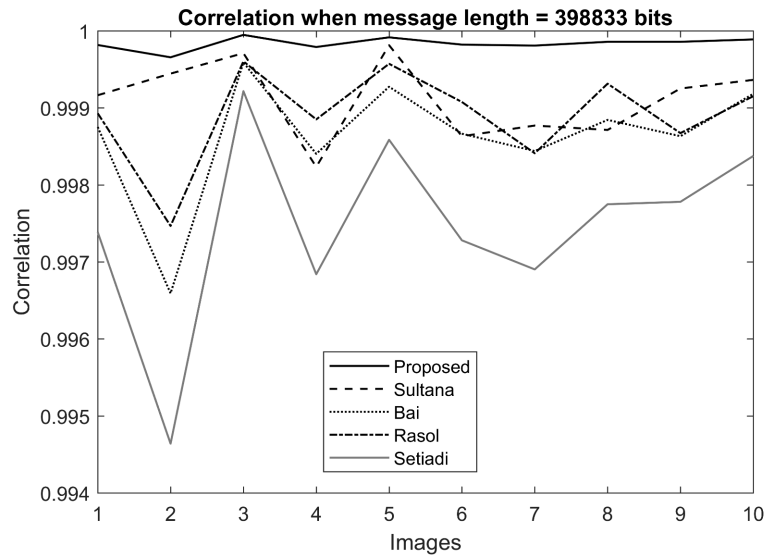


Figure 7: Performance comparison of the proposed scheme with Sultana [32], Bai [5], Rasol [33], and Setiadi [31] in terms of Correlation coefficients when message length = 398833 bits.

Table 8: PSNR values of different schemes

Methods	ImageName	Message length (bits)		
		274661	442483	693637
Proposed	F16.jpg	62.47	51.19	42.69
	boat.jpg	54.96	46.28	41.79
	pepper.jpg	57.92	50.40	46.89
[32]	F16.jpg	53.12	52.78	52.78
	boat.jpg	48.00	46.88	46.88
	pepper.jpg	52.58	51.08	51.08
[5]	F16.jpg	52.95	52.23	52.23
	boat.jpg	48.16	47.20	47.20
	pepper.jpg	52.17	50.96	50.96
[33]	F16.jpg	53.63	52.93	52.93
	boat.jpg	48.73	47.93	47.93
	pepper.jpg	55.10	54.37	54.37
[31]	F16.jpg	46.37	44.16	43.91
	boat.jpg	46.11	43.73	43.15
	pepper.jpg	50.69	48.13	47.51

Table 9 contains the time complexity for message lengths 274661 bits, 442483 bits, and 693637 bits.

**Difference of standard Deviation between stego and cover when message length = 398833 bits**

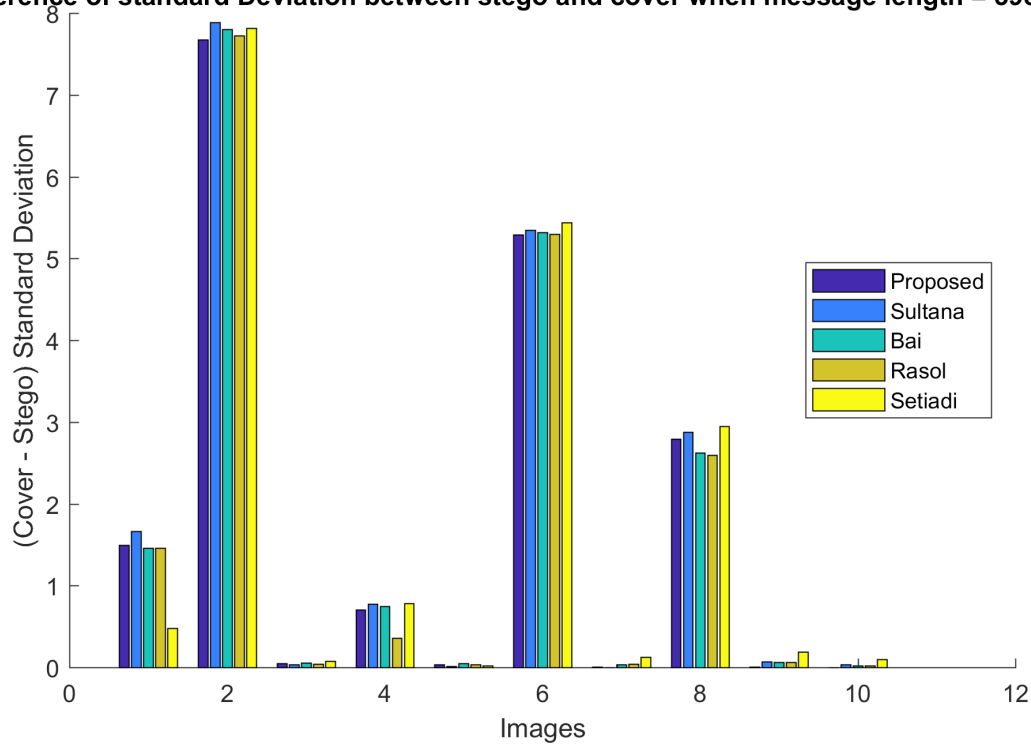


Figure 8: Difference of standard Deviations of cover and stego images when message length = 398833 bits. The figure states that the differences are very small and close to each other.

Table 9: Elapsed time of different schemes

Methods	ImageName	Message length (bits)		
		274661	442483	693637
Proposed method	F16.jpg	13.18	13.48	14.12
	boat.jpg	12.95	13.58	15.07
	pepper.jpg	12.90	14.16	14.69
[32]	F16.jpg	11.72	13.69	13.27
	boat.jpg	10.29	13.33	13.38
	pepper.jpg	10.51	14.13	13.68
[5]	F16.jpg	10.86	13.41	13.54
	boat.jpg	10.68	13.42	13.33
	pepper.jpg	11.17	13.65	13.59
[33]	F16.jpg	11.47	13.52	13.35
	boat.jpg	11.21	13.52	13.38
	pepper.jpg	11.53	13.42	13.32
[31]	F16.jpg	8.59	13.36	13.91
	boat.jpg	7.66	12.04	13.88
	pepper.jpg	7.23	12.78	14.25

Table 10: Correlation values of different schemes

Methods	ImageName	Message length (bits)		
		274661	442483	693637
Proposed method	F16.jpg	0.999	0.998	0.996
	boat.jpg	0.999	0.997	0.995
	pepper.jpg	0.999	0.998	0.996
[32]	F16.jpg	0.999	0.999	0.999
	boat.jpg	0.998	0.998	0.998
	pepper.jpg	0.999	0.998	0.998
[5]	F16.jpg	0.998	0.998	0.998
	boat.jpg	0.998	0.998	0.998
	pepper.jpg	0.999	0.998	0.998
[33]	F16.jpg	0.999	0.998	0.998
	boat.jpg	0.999	0.998	0.998
	pepper.jpg	0.999	0.999	0.999
[31]	F16.jpg	0.998	0.997	0.996
	boat.jpg	0.997	0.996	0.996
	pepper.jpg	0.998	0.997	0.997

Table 11: Standard deviations values of different schemes

Methods	ImageName	Message length (bits)		
		274661	442483	693637
Proposed method	F16.jpg	1.49	1.45	0.38
	boat.jpg	0.71	0.76	0.50
	pepper.jpg	2.58	2.65	2.77
[32]	F16.jpg	1.65	1.66	1.66
	boat.jpg	0.74	0.77	0.77
	pepper.jpg	2.86	2.87	2.87
[5]	F16.jpg	1.46	1.46	1.46
	boat.jpg	0.72	0.74	0.74
	pepper.jpg	2.61	2.62	2.62
[33]	F16.jpg	1.46	1.46	1.46
	boat.jpg	0.34	0.35	0.35
	pepper.jpg	2.59	2.59	2.59
[31]	F16.jpg	0.53	0.47	0.47
	boat.jpg	0.85	0.80	0.83
	pepper.jpg	2.93	2.93	2.97

## 5 Robustness of the proposed scheme against attacks

We statistically analyzed our scheme using various parameters such as correlation coefficient, standard deviation, entropy, cosine similarities, and pixel difference histogram to check its robustness against various attacks. We first measured correlation coefficients  $\rho_s C$  between the cover and stego image for message length 398833 bits.  $\rho_s C = 0$  stands for no relationship between two images.  $\rho_s C > 0$  means a positive correlation between the cover and stego image and lies perfect relationship when it reaches 1. Similarly, a negative value of  $\rho_s C$  indicates a negative relationship. Results of  $\rho_s C$  are depicted in Figure 7. Through proposed method shows a higher correlation value, its difference from others is insignificance.

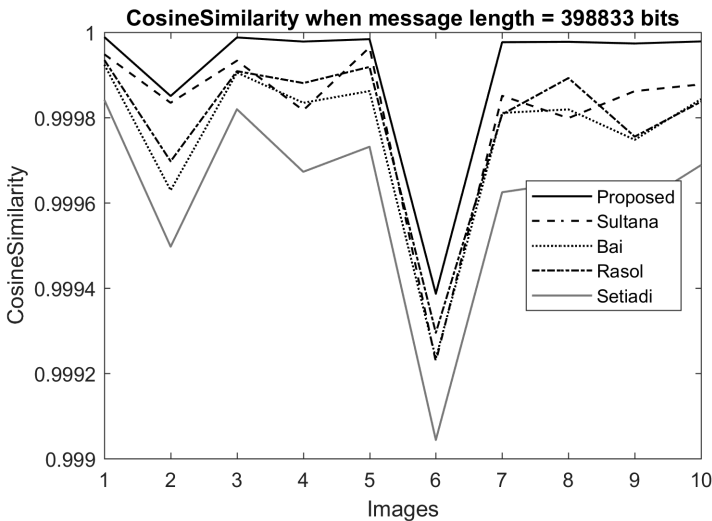


Figure 9: Performance comparison of the proposed scheme with Sultana [32], Bai [5], Rasol [33], and Setiadi [31] in terms of Cosine similarity when message length = 398833 bits.

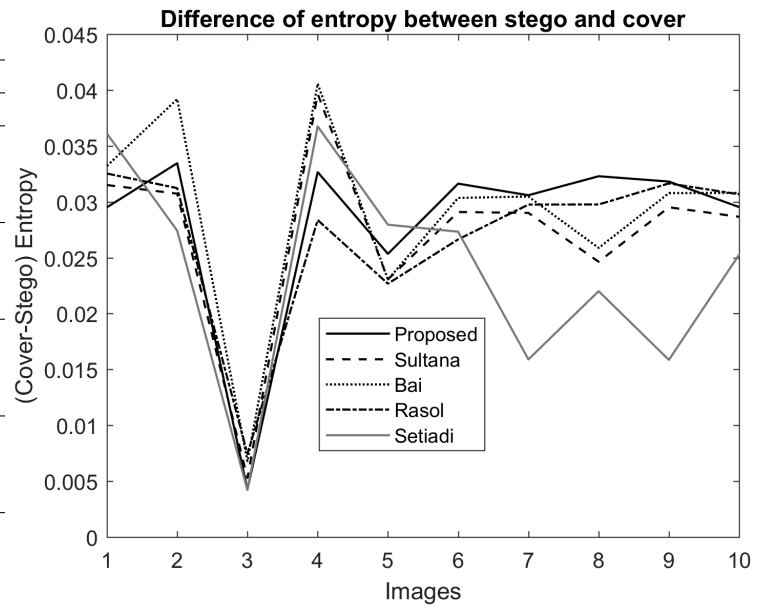


Figure 10: Performance comparison of the proposed scheme with Sultana [32], Bai [5], Rasol [33], and Setiadi [31] in terms of entropy values when message length = 398833 bits.

Table 12: Cosine similarity values of different schemes

Methods	ImageName	Message length (bits)		
		274661	442483	693637
Proposed method	F16.jpg	0.9999	0.9998	0.9997
	boat.jpg	0.9999	0.9997	0.9995
	pepper.jpg	0.9999	0.9997	0.9995
[32]	F16.jpg	0.9999	0.9999	0.9999
	boat.jpg	0.9998	0.9998	0.9998
	pepper.jpg	0.9998	0.9997	0.9997
[5]	F16.jpg	0.9999	0.9999	0.9999
	boat.jpg	0.9998	0.9998	0.9998
	pepper.jpg	0.9998	0.9998	0.9998
[33]	F16.jpg	0.9999	0.9999	0.9999
	boat.jpg	0.9999	0.9998	0.9998
	pepper.jpg	0.9999	0.9998	0.9998
[31]	F16.jpg	0.9998	0.9998	0.9998
	boat.jpg	0.9997	0.9996	0.9995
	pepper.jpg	0.9997	0.9996	0.9995

Rather, as with others, it represents a higher correlation between cover and stego image. We also measured the standard deviation of pixel values from, their mean, separately in cover and stego image is  $\sigma_c$  and  $\sigma_s$  for message length 398833 bits. We then calculated their difference by  $\sigma_d = \sigma_c - \sigma_s$ . Ideally,  $\sigma_d$  should be zero for a non-tempered image. The results are shown in Figure 8.

Table 10 and Table 11 contains the values of correlation coefficients and standard deviations for message length 274661 bits, 442483 bits, and 693637 bits.

To verify further with similar statistics, we measured cosine similarities between the cover and stego images for a message length of 398833 bits. That value is 1 for two identical images and 0 for two fully mismatched images. The results are demonstrated in Figure 9.

The figure illustrates that our proposed method shows higher values than the other competing schemes.

We computed, the entropy values  $H$  as well in cover and stego images for message length 398833 bits. Next, we calculated their differences. That difference value is zero for two identical images. Results are plotted in Figure 10. The figure shows that none of the results are greater than 0.04, i.e., these are very small and close to zero.

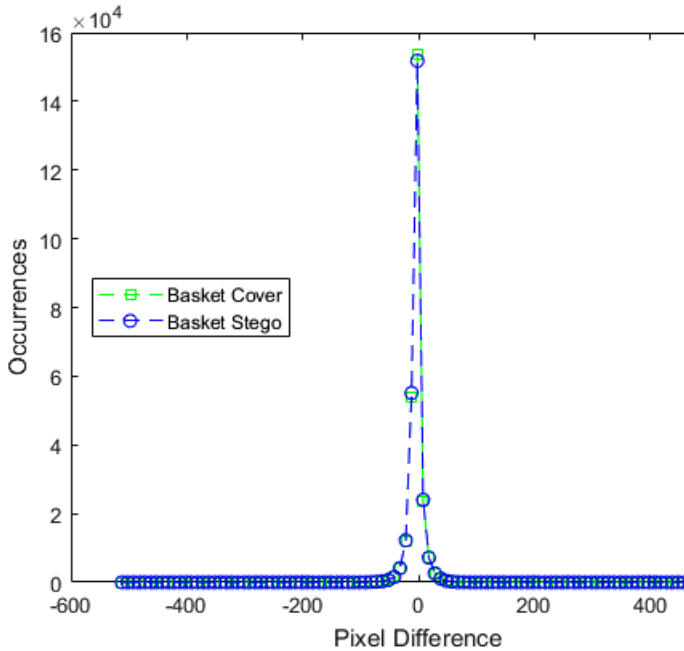


Figure 11: Performance comparison of the proposed scheme with Sultana, Bai, Rasol, and Setiadi in terms of pixel difference histogram when message length = 398833 bits for basket cover image.

Table 12 and Table 13 contains the values of cosine similarity and entropy for message length 274661 bits, 442483 bits, and 693637 bits.

Table 13: Entropy values of different schemes

Methods	ImageName	Message length (bits)		
		274661	442483	693637
Proposed method	F16.jpg	0.027	0.040	0.052
	boat.jpg	0.030	0.046	0.068
	pepper.jpg	0.028	0.029	0.046
[32]	F16.jpg	0.024	0.031	0.031
	boat.jpg	0.020	0.039	0.039
	pepper.jpg	0.011	0.024	0.024
[5]	F16.jpg	0.022	0.033	0.033
	boat.jpg	0.023	0.040	0.040
	pepper.jpg	0.014	0.025	0.025
[33]	F16.jpg	0.022	0.032	0.032
	boat.jpg	0.019	0.028	0.028
	pepper.jpg	0.021	0.029	0.029
[32]	F16.jpg	0.021	0.043	0.048
	boat.jpg	0.019	0.048	0.069
	pepper.jpg	0.0008	0.032	0.044

We also used, another statistical tool pixel difference histogram (PDH) to identify the stego images. The PDH of the original images and corresponding stego images are shown in Figure 11 and Figure 12 for message length 398833 bits.

Thus, it can be deduced from the results of these experiments that our method is strong enough to protect against attacks on implanted data.

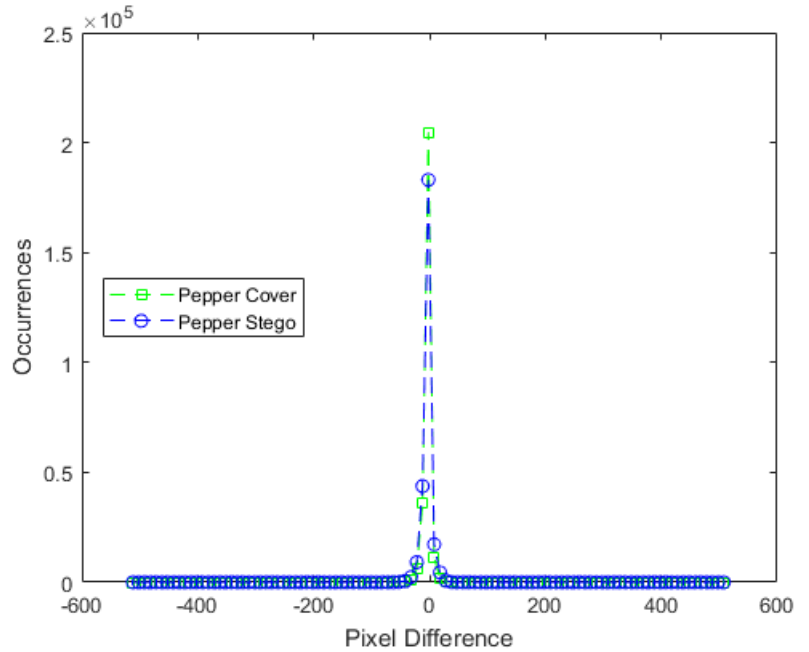


Figure 12: Performance comparison of the proposed scheme with Sultana, Bai, Rasol, and Setiadi in terms of pixel difference histogram when message length = 398833 bits for peppers cover image.

## 6 Conclusion

The edge-detection-based embedding schemes, generally, implant either in edge pixels only or a different number of bits in edge and non-edge pixels. Few of them use multiple edge detectors and hybridize them to increase the number of edge pixels in the resultant edge image. Even, either maintaining the visual quality of stego image or meeting the embedding demand is still functioning as a challenging matter. This research overcomes that problem by realizing the situation and then dynamically choosing the number of implanted bits per pixel, the required edge detectors, and the best hybridization technique.

Edge detectors are selected and their hybridization is performed based on message length. This scheme, thus, increases or decreases the number of edge pixels and fixes the number of implanted bits per edge and non-edge pixels according to the embedding demand. The experimental result deduces that this scheme embeds 92.37%, 73.92%, 74.78%, and 9.60% more bits than [32], [5], [33], [31], respectively. Similarly, for a small embedding demand, the proposed scheme demonstrates 37.45%, 46.87%, 44.21%, and 55.56% higher PSNR values than [32], [5], [33], [31], respectively. Moreover, the statistical analyses state that this scheme exhibits stronger security against attacks.

In our future work, we wish to apply this high-performing embedding scheme in electronic kit development for use in healthcare and forensic-related applications. We also plan to apply a machine learning algorithm for segmenting the image areas into different complex levels so that we can implant a different number of bits in those staged areas, i.e. levels.

**Conflict of Interest** There is nothing related to conflict of interest.

**Acknowledgment** The first author is a research fellow of the ICT division of the Ministry of Post, Telecommunication, and Information Technology of the Government of Bangladesh. Therefore, we want to devote a thank to the concerned ministry, and at the same time, we would like to acknowledge that support.

## References

- [1] C. Vanmathi, S. Prabu, "Image steganography using fuzzy logic and chaotic for large payload and high imperceptibility," *International Journal of Fuzzy Systems*, **20**(2), 460–473, 2018, doi:10.1007/s40815-017-0420-0.
- [2] E. J. Kusuma, O. R. Indriani, C. A. Sari, E. H. Rachmawanto, et al., "An imperceptible LSB image hiding on edge region using DES encryption," in 2017 International Conference on Innovative and Creative Information Technology (ICITech), 1–6, IEEE, 2017, doi:10.1109/INNOCIT.2017.8319132.
- [3] M. Hussain, A. W. A. Wahab, N. Javed, K.-H. Jung, "Recursive information hiding scheme through LSB, PVD shift, and MPE," *IETE Technical Review*, **35**(1), 53–63, 2018, doi:10.1080/02564602.2016.1244496.
- [4] W.-J. Chen, C.-C. Chang, T. H. N. Le, "High payload steganography mechanism using hybrid edge detector," *Expert Systems with applications*, **37**(4), 3292–3301, 2010, doi:10.1016/j.eswa.2009.09.050.
- [5] J. Bai, C.-C. Chang, T.-S. Nguyen, C. Zhu, Y. Liu, "A high payload steganographic algorithm based on edge detection," *Displays*, **46**, 42–51, 2017, doi:10.1016/j.displa.2016.12.004.
- [6] H. Sultana, A. Kamal, "An Edge Detection Based Reversible Data Hiding Scheme," in 2022 IEEE Delhi Section Conference (DELCON), 1–6, IEEE, 2022, doi:10.1109/DELCON54057.2022.9753404.
- [7] P. W. Adi, F. Z. Rahmanti, N. A. Abu, "High quality image steganography on integer Haar Wavelet Transform using modulus function," in 2015 International Conference on Science in Information Technology (ICSITech), 79–84, IEEE, 2015, doi:10.1109/ICSITech.2015.7407781.
- [8] H. Al-Dmour, A. Al-Ani, "A steganography embedding method based on edge identification and XOR coding," *Expert systems with Applications*, **46**, 293–306, 2016, doi:10.1016/j.eswa.2015.10.024.
- [9] A. Kamal, M. M. Islam, "A prediction error based histogram association and mapping technique for data embedment," *Journal of Information Security and Applications*, **48**, 102368, 2019, doi:10.1016/j.jisa.2019.102368.
- [10] H. Sultana, A. Kamal, M. M. Islam, "Enhancing the Robustness of Visual Degradation Based HAM Reversible Data Hiding," *J. Comput. Sci.*, **12**(2), 88–97, 2016, doi:10.3844/jcssp.2016.88.97.
- [11] A. Kamal, M. M. Islam, "Enhancing embedding capacity and stego image quality by employing multi predictors," *Journal of Information Security and Applications*, **32**, 59–74, 2017, doi:10.1016/j.jisa.2016.08.005.
- [12] A. Kamal, M. M. Islam, "Boosting up the data hiding rate through multi cycle embedment process," *Journal of Visual Communication and Image Representation*, **40**, 574–588, 2016, doi:10.1016/j.jvcir.2016.07.023.
- [13] A. Kamal, M. M. Islam, "Capacity improvement of reversible data hiding scheme through better prediction and double cycle embedding process," in 2015 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), 1–6, IEEE, 2015, doi:10.1109/ANTS.2015.7413636.
- [14] A. Kamal, M. M. Islam, "An image distortion-based enhanced embedding scheme," *Iran Journal of Computer Science*, **1**(3), 175–186, 2018, doi:10.1007/s42044-018-0016-3.
- [15] A. Kamal, M. M. Islam, Z. Islam, "An embedding technique for smartcard-supported e-healthcare services," *Iran Journal of Computer Science*, **3**(4), 195–205, 2020, doi:10.1007/s42044-020-00055-1.
- [16] A. Kamal, M. M. Islam, "Enhancing the embedding payload by handling the affair of association and mapping of block pixels through prediction errors histogram," in 2016 International Conference on Networking Systems and Security (NSysS), 1–8, IEEE, 2016, doi:10.1109/NSysS.2016.7400691.
- [17] A. Kamal, M. M. Islam, "Enhancing the performance of the data embedment process through encoding errors," *Electronics*, **5**(4), 79, 2016, doi:10.3390/electronics5040079.
- [18] A. Kamal, M. M. Islam, "Facilitating and securing offline e-medicine service through image steganography," *Healthcare Technology Letters*, **1**(2), 74–79, 2014, doi:10.1049/htl.2013.0026.
- [19] W. Hong, "Adaptive reversible data hiding method based on error energy control and histogram shifting," *Optics Communications*, **285**(2), 101–108, 2012, doi:10.1016/j.optcom.2011.09.005.
- [20] H. Yao, C. Qin, Z. Tang, Y. Tian, "Improved dual-image reversible data hiding method using the selection strategy of shiftable pixels' coordinates with minimum distortion," *Signal Processing*, **135**, 26–35, 2017, doi:10.1016/j.sigpro.2016.12.029.
- [21] S. Yi, Y. Zhou, "Binary-block embedding for reversible data hiding in encrypted images," *Signal Processing*, **133**, 40–51, 2017, doi:10.1016/j.sigpro.2016.10.017.
- [22] S. Khan, N. Ahmad, M. Ismail, N. Minallah, T. Khan, "A secure true edge based 4 least significant bits steganography," in 2015 International Conference on Emerging Technologies (ICET), 1–4, IEEE, 2015, doi:10.1109/ICET.2015.7389227.
- [23] G. Swain, "Adaptive pixel value differencing steganography using both vertical and horizontal edges," *Multimedia Tools and Applications*, **75**(21), 13541–13556, 2016, doi:10.1007/s11042-015-2937-2.
- [24] S. Sun, "A novel edge based image steganography with 2k correction and Huffman encoding," *Information Processing Letters*, **116**(2), 93–99, 2016, doi:10.1016/j.ipl.2015.09.016.
- [25] C.-F. Lee, C.-Y. Weng, K.-C. Chen, "An efficient reversible data hiding with reduplicated exploiting modification direction using image interpolation and edge detection," *Multimedia Tools and Applications*, **76**(7), 9993–10016, 2017, doi:10.1007/s11042-016-3591-z.
- [26] H.-W. Tseng, H.-S. Leng, "High-payload block-based data hiding scheme using hybrid edge detector with minimal distortion," *IET Image Processing*, **8**(11), 647–654, 2014, doi:10.1049/iet-ipr.2013.0584.
- [27] S. Kumar, A. Singh, M. Kumar, "Information hiding with adaptive steganography based on novel fuzzy edge identification," *Defence Technology*, **15**(2), 162–169, 2019, doi:10.1016/j.dt.2018.08.003.
- [28] C.-F. Lee, C.-C. Chang, X. Xie, K. Mao, R.-H. Shi, "An adaptive high-fidelity steganographic scheme using edge detection and hybrid hamming codes," *Displays*, **53**, 30–39, 2018, doi:10.1016/j.displa.2018.06.001.
- [29] S. K. Ghosal, A. Chatterjee, R. Sarkar, "Image steganography based on Kirsch edge detection," *Multimedia Systems*, **27**(1), 73–87, 2021, doi:10.1007/s00530-020-00703-3.
- [30] B. Vishnu, L. V. Namboothiri, S. R. Sajeesh, "Enhanced image steganography with PVD and edge detection," in 2020 Fourth International Conference on Computing Methodologies and Communication (ICCCMC), 827–832, IEEE, 2020, doi:10.1109/ICCCMC48092.2020.ICCCMC-000153.
- [31] D. R. I. M. Setiadi, "Improved payload capacity in LSB image steganography uses dilated hybrid edge detection," 2022, doi:10.1016/j.jksuci.2019.12.007.

- [32] H. Sultana, A. Kamal, "Image Steganography System based on Hybrid Edge Detector," in 2021 24th International Conference on Computer and Information Technology (ICCIT), 1–6, IEEE, 2021, doi:[10.1109/ICCIT54785.2021.9689777](https://doi.org/10.1109/ICCIT54785.2021.9689777).
- [33] D. R. I. M. Setiadi, J. Jumanto, "An enhanced LSB-image steganography using the hybrid Canny-Sobel edge detection," *Cybernetics and Information Technologies*, **18**(2), 74–88, 2018, doi:[10.2478/cait-2018-0029](https://doi.org/10.2478/cait-2018-0029).
- [34] K. M. Sagayam, A. A. Anton Jone, K. Cengiz, L. Rajesh, A. A. Elngar, "Breast Cancer Detection based on 3-D Mammography Images using Deep Learning Strategies," *Journal of Information Technology Management*, **14**(4), 2–18, 2022, doi:[10.22059/JITM.2022.88132](https://doi.org/10.22059/JITM.2022.88132).
- [35] P. Elayaraja, S. Kumarganesh, K. Martin Sagayam, H. Dang, M. Pomplun, "An efficient approach for detection and classification of cancer regions in cervical images using optimization based CNN classification approach," *Journal of Intelligent & Fuzzy Systems*, (Preprint), 1–11, 2022, doi:[10.3233/JIFS-212871](https://doi.org/10.3233/JIFS-212871).
- [36] R. Jayasingh, J. K. RJS, D. B. Telagathoti, K. M. Sagayam, S. Pramanik, O. P. Jena, S. K. Bandyopadhyay, "Speckle noise removal by SORAMA segmentation in Digital Image Processing to facilitate precise robotic surgery," *International Journal of Reliable and Quality E-Healthcare (IJRQEH)*, **11**(1), 1–19, 2022, doi:[10.4018/IJRQEH.295083](https://doi.org/10.4018/IJRQEH.295083).
- [37] T. B. Mary, P. M. Bruntha, M. Manimekalai, K. M. Sagayam, H. Dang, et al., "Investigation of an Efficient Integrated Semantic Interactive Algorithm for Image Retrieval," *Pattern Recognition and Image Analysis*, **31**(4), 709–721, 2021, doi:[10.1134/S1054661821040234](https://doi.org/10.1134/S1054661821040234).
- [38] A. D. Andrushia, K. M. Sagayam, H. Dang, M. Pomplun, L. Quach, "Visual-Saliency-Based Abnormality Detection for MRI Brain Images—Alzheimer's Disease Analysis," *Applied Sciences*, **11**(19), 9199, 2021, doi:[10.3390/app11199199](https://doi.org/10.3390/app11199199).
- [39] J. Andrew, T. Mhatesh, R. D. Sebastin, K. M. Sagayam, J. Eunice, M. Pomplun, H. Dang, "Super-resolution reconstruction of brain magnetic resonance images via lightweight autoencoder," *Informatics in Medicine Unlocked*, **26**, 100713, 2021, doi:[10.1016/j.imu.2021.100713](https://doi.org/10.1016/j.imu.2021.100713).
- [40] J. A. Onesimu, J. Karthikeyan, D. S. J. Viswas, R. D. Sebastian, "Security and Privacy Challenges of Deep Learning: A Comprehensive Survey," *Research Anthology on Privatizing and Securing Data*, 1258–1280, 2021, doi:[10.4018/978-1-7998-8954-0.ch059](https://doi.org/10.4018/978-1-7998-8954-0.ch059).
- [41] J. A. Onesimu, J. Karthikeyan, Y. Sei, "An efficient clustering-based anonymization scheme for privacy-preserving data collection in IoT based healthcare services," *Peer-to-Peer Networking and Applications*, **14**(3), 1629–1649, 2021, doi:[10.1007/s12083-021-01077-7](https://doi.org/10.1007/s12083-021-01077-7).
- [42] Y. Sei, J. Andrew, H. Okumura, A. Ohsuga, "Privacy-Preserving Collaborative Data Collection and Analysis with Many Missing Values," *IEEE Transactions on Dependable and Secure Computing*, 2022, doi:[10.1109/TDSC.2022.3174887](https://doi.org/10.1109/TDSC.2022.3174887).
- [43] A. Jan, S. A. Parah, M. Hassan, B. A. Malik, "Realization of Efficient Steganographic Scheme Using Hybrid Edge Detection and Chaos," *Arabian Journal for Science and Engineering*, 1–14, 2022, doi:[10.1007/s13369-022-06960-w](https://doi.org/10.1007/s13369-022-06960-w).

# On the Polytopic Modelling & Robust $H_\infty$ Control of Nonlinear Systems Subject to Cyber-attack: Application to Attitude Stabilization of Quadrotor

Bezzaoucha-Rebaï Souad \*

EIGSI- La Rochelle, 17041, France

## ARTICLE INFO

Article history:

Received: 05 November, 2022

Accepted: 08 January, 2023

Online: 24 January, 2023

Keywords:

Quadrotor

Stabilization

Polytopic representation

Stealthy attacks

## ABSTRACT

In the present contribution, a robust output  $H_\infty$  control ensuring the stability, reliability and security for nonlinear systems when actuator attacks (data deception attacks) occur. A new design method based on the polytopic rewriting of the attacked system as an uncertain one subject to external disturbances will be detailed. Robust polytopic state feedback observer stabilizing controller based on the PDC (Parallel Distributed Compensation) polytopic framework with disturbance attenuation for the obtained uncertain system will also be considered. The obtained methodology is used to ensure the stability and security of a quadrotor/UAV subject to stealthy actuator attacks. State and attacks estimations signals are given in order to highlight the efficiency of the developed approach.

## 1 Introduction

Based on previous contribution [1], this paper is an extension of the original work which aims to ensure a robust attitude stabilization of a quadrotor subject to stealthy actuator attacks. The modelling and control aspect were considered in this first contribution, where in the following the observer design for both state and stealthy attacks is added. Robust polytopic state feedback stabilizing controller based on PDC (Parallel Distributed Compensation) polytopic framework observer with disturbance attenuation (guaranteed by the  $H_\infty$  norm) for the obtained uncertain system is also considered.

Design and implementing feedback control strategies that are robust against cyber-attacks is of critical importance nowadays. Assuming that the behavior of the system is driven via actuator commands, the actuator data deception attack corresponds to a manipulation of an attacker on the communication channels between the plant and the controller. The actuator commands are then corrupted and it becomes necessary to integrate this data in the control system design and make it as robust as possible to these stealthy attacks.

The objective is then the design of a resilient control for a system where an attacker corrupts control packets; and of course, to detect and reduce/attenuate the effect of these corrupted signals on the well-behaviour of the considered system.

In the following contribution, the novelty comparing the work

originally presented in [1] is about the estimation and robust control part where both state and stealthy attacks are now estimated with an  $H_\infty$  disturbance attenuation property.

One solution in order to represent and implement heuristic knowledge to control nonlinear systems when remaining the study relatively simple consists in the use of the polytopic Takagi-Sugeno (T-S) structure. Indeed, this representation was initially proposed by [2] and [3], and proven its efficiency in various applications in the past decades [4].

Based on the polytopic T-S approach, a number of most important issues in control systems have been addressed in the past few years. These includes stability analysis [5], state and output feedback control [6], [7], performances and robustness [8], [4], as well as recently cyber-security [9]–[12].

Solving the considered problem ( $H_\infty$  control of stabilization), the nonlinear behaviour of the quadrotor, including stealthy attacks, is represented in terms of an uncertain polytopic system subject to bounded external disturbances. The stabilization and robust  $H_\infty$  control, based on state estimation feedback is then deduced based on classical Lyapunov theory leading to a set of matrices inequalities (constraints) to solve. These constraints, solvable through convex optimization techniques allows to obtain polytopic controllers and observers that guarantee both stability and robustness of the closed-loop system.

Indeed, this paper focuses on the problem of observer-based

\*Corresponding Author: BEZZAOUCHA-REBAÏ, Souad. EIGSI-La Rochelle; 26 Rue François de Vaux de Foletier, 17000, France. Email: souad.bezzaoucha@eigsi.fr



$H_\infty$  control for polytopic T-S systems under actuator data deception attacks. Sufficient conditions for the simultaneous controller and observer design with a desired  $H_\infty$  disturbance attenuation level are derived in terms of linear matrix inequalities which can be easily solved by using available software package (Matlab for example).

The present paper objective is to contribute to the cybersecurity and resilience design of Unmanned Aerial Vehicles (UAVs). It is known that the wireless control used to monitor drones makes them defenceless to a large variety of cyberattacks, which may have severe consequences on the system behaviour/security/integrity/performances.

In this contribution, stealthy attacks disturbing and destabilizing the control and navigation system of the UAV are studied. We aim to propose a robust control ensuring the safety and security of the UAV despite these assaults.

This contribution strategy was previously developed for cyberattacks estimation [1], [10] and is now adapted to the quadrotor robust control. The estimation of the system states and stealthy signals will be given.

Considering the estimable premise variables, the attacked system will be presented as an uncertain T-S model. Based on the resulting system, a PDC observer based control will be designed.

The paper organized as follows: a brief state of the art is presented in section 1; the system modelling with the actuator data deception attack and the uncertain system representation are detailed respectively in section 2 and 3. Section 4 is about the robust output  $H_\infty$  observe-based T-S controller. Section 5 is about the approach illustration through an application to quadrotor attitude stabilization with simulation results. The final section, 6 is about conclusion and perspectives.

The applied methodology solving the considered problem is the following:

1. The modelling aspect: such that the nonlinear behaviour and threats attacks are both represented in a polytopic T-S form based on the sector nonlinearity transformation (SNT); it is important to note that in this representation, there is no approximation or any loss of information. The main advantage of this method consists into an exact rewriting of the original nonlinear equations.
2. In order to be able to implement the state feedback observer based PDC  $H_\infty$  control law, the main constraint in the obtained model (5) is about the immeasurable state and time-varying parameters present in the weighting functions  $h_i$  &  $\mu_j$ , and the control law. In order to overcome this difficulty, it is imperative to standardize the system equations in order to have only measurable and/or estimable premise variables. For that, an uncertain representation of the system equations (5) is proposed in section 3 in order to obtain a more convenient model for the study; i.e. (19).
3. Based on the chosen structure for the observer and control law, a robust  $H_\infty$  T-S control of the nonlinear system is considered in section 4. The objective, in addition to the system stabilization, is to ensure an attenuation of the external perturbation, guaranteed thanks to the  $H_\infty$  norm.

4. The final step of our study would be the application of the developed approach to our case study; i.e. the attitude stabilization of the quadrotor.

## 2 System Modelling: a Polytopic representation

In the following section, based on the nonlinear state space model of a system, a polytopic representation will be deduced applying the sector nonlinearity approach (SNT).

Assuming that our system is subject to actuator data deception attacks (modeled as unknown, but bounded, multiplicative time-varying parameters); our nonlinear model under these attacks may be represented by the following state space system equations:

$$\begin{cases} \dot{x}(t) &= \sum_{i=1}^r h_i(t)(A_i x(t) + B_i(t)u(t)) \\ y(t) &= Cx(t) \end{cases} \quad (1)$$

s.t. matrices  $B_i(t)$  are defined as:

$$B_i(t) := B_i + \Gamma^u a^u(t) \quad (2)$$

In this model,  $B_i$  is called the nominal input matrix (i.e. when none attack occurring);  $\Gamma^u$  is known as the binary incidence matrix, which indicates the data channels that can be accessed by the attacker; finally  $a^u(t)$  represents the actuator data corruption signal.

The stealthy signals  $a^u(t)$  are unknown (in terms of value), but bounded (limits assumed to be known)  $a^u(t) \in [a_2^u, a_1^u]$ . Applying the SNT transformation, the following representation is proposed:

$$a^u(t) = \mu_1(a^u(t))a_1^u + \mu_2(a^u(t))a_2^u \quad (3)$$

with

$$\begin{cases} \mu_1(a^u(t)) &= \frac{a^u(t) - a_2^u}{a_1^u - a_2^u} \\ \mu_2(a^u(t)) &= \frac{a_1^u - a^u(t)}{a_1^u - a_2^u} \end{cases} \quad (4)$$

$$\mu_1(a^u(t)) + \mu_2(a^u(t)) = 1, \quad \forall t$$

The equations (1) of the attacked system is then rewritten as:

$$\begin{cases} \dot{x}(t) &= \sum_{i=1}^r \sum_{j=1}^2 h_i(t)\mu_j(t)(A_i x(t) + \mathcal{B}_{ij}u(t)) \\ y(t) &= Cx(t) \end{cases} \quad (5)$$

s.t.  $\mathcal{B}_i^j(t)$  are defined as:

$$\mathcal{B}_{i1} = B_i + a_1^u \Gamma^u, \quad \mathcal{B}_{i2} = B_i + a_2^u \Gamma^u; \quad i = 1, 2 \quad (6)$$

## 3 Uncertain System Representation

Based on contributions [11] and [12], where the data deception representation of cyber-attacks via a time-varying model and thanks to a polytopic form of an uncertain system representation, a state

and actuator data deception attacks observer is proposed and given by the following equations:

$$\begin{cases} \dot{\hat{x}}(t) = \sum_{i=1}^r \sum_{j=1}^2 h_i(\hat{x}(t)) \mu_j(\hat{a}^u(t)) \\ \quad (A_i \hat{x}(t) + \mathcal{B}_{ij} u(t) + L_{ij}(y(t) - \hat{y}(t))) \\ \dot{\hat{a}}^u(t) = \sum_{i=1}^r \sum_{j=1}^2 h_i(\hat{x}(t)) \mu_j(\hat{a}^u(t)) \\ \quad (K_{ij}(y(t) - \hat{y}(t)) - \alpha_{ij}^u \hat{a}^u(t)) \\ \hat{y}(t) = C \hat{x}(t) \end{cases} \quad (7)$$

s.t.  $L_{ij} \in \mathbb{R}^{n_x \times m}$ ,  $K_{ij}^u \in \mathbb{R}^{n \times m}$  and  $\alpha_{ij}^u \in \mathbb{R}^{n \times n}$  are solution of a LMI- $H_{\infty 2}$  attenuation conditions ensuring both estimation errors for the states and malicious input parameters to converge to zero. Let us now define the estimation errors  $e_x(t)$  and  $e_{a^u}(t)$  (for the state and cyber-attacks) as:

$$\begin{aligned} e_x(t) &= x(t) - \hat{x}(t) \\ e_{a^u}(t) &= a^u(t) - \hat{a}^u(t) \end{aligned} \quad (8)$$

The system equations (5) can be rewritten as follows [13], [11]:

$$\begin{cases} \dot{x}(t) = \sum_{i=1}^r \sum_{j=1}^2 [h_i(\hat{x}(t)) \mu_j(\hat{a}^u(t)) (A_i x(t) + \mathcal{B}_{ij} u(t)) + \\ \quad \delta_{ij}(t) (A_i x(t) + \mathcal{B}_{ij} u(t))] \\ y(t) = C x(t) \end{cases} \quad (9)$$

with  $\delta_{ij}(t)$  are defined by the following equations:

$$\delta_{ij}(t) = h_i(x(t)) \mu_j(a^u(t)) - \mu_j(\hat{x}(t)) \mu_j(\hat{a}^u(t)) \quad (10)$$

satisfying:

$$-1 \leq \delta_{ij}(t) \leq 1 \quad (11)$$

Let us introduce now:

$$\Delta A(t) = \sum_{i=1}^r \sum_{j=1}^2 \delta_{ij}(t) A_i = \mathcal{A} \Sigma(t) E_A \quad (12)$$

$$\Delta B(t) = \sum_{i=1}^r \sum_{j=1}^2 \delta_{ij}(t) \mathcal{B}_{ij} = \mathcal{B} \Sigma(t) E_B \quad (13)$$

with

$$\mathcal{A} = \left[ \underbrace{A_1 \quad A_1}_{2 \text{ times}} \quad \dots \quad \underbrace{A_r \quad \dots \quad A_r}_{2 \text{ times}} \right] \quad (14)$$

$$\mathcal{B} = \left[ \mathcal{B}_1^1 \quad \dots \quad \mathcal{B}_r^2 \right] \quad (15)$$

$$\Sigma(t) = \text{diag}(\delta_{11}(t), \dots, \delta_{r2}(t)), \quad (16)$$

$$E_A = \left[ I_{n_x} \quad \dots \quad I_{n_x} \right]^T, \quad E_B = \left[ I_{n_u} \quad \dots \quad I_{n_u} \right]^T \quad (17)$$

Thanks to (11) and definitions (16), we have:

$$\Sigma^T(t) \Sigma(t) \leq I \quad (18)$$

Using the above definitions (12)-(17), system (11) is then written as an uncertain system given by:

$$\begin{cases} \dot{x}(t) = \sum_{i=1}^r \sum_{j=1}^2 h_i(\hat{x}(t)) \mu_j(\hat{a}^u(t)) \\ \quad ((A_i + \Delta A(t)) x(t) + (\mathcal{B}_{ij} + \Delta B(t)) u(t)) \\ y(t) = C x(t) \end{cases} \quad (19)$$

## 4 Robust Output $H_{\infty}$ T-S Control

The objective in the following section is to determine polytopic T-S controller and observer gains ensuring that:

- The system given by (19) is asymptotically stable in the presence of data deception attacks.
- The attenuation of the external perturbations like (i.e. actuator attacks) is guaranteed by the  $H_{\infty}$  norm. i.e. find for a given scalar  $\gamma > 0$ , an observer (7) and a PDC controller (20) such that the attenuation condition (26) is satisfied. The resulting conditions to be solved will be given in Lemma 2.

Considering the nonlinear system subject to data deception attacks given by the system equations (1), and the polytopic T-S observer to estimate the unmeasurable state variables and unknown time-varying parameter (actuator attack signal  $a^u(t)$ ) given by system equations (7) with the following PDC (Parallel Distributed Compensation) controller defined by:

$$u(t) = - \sum_{k=1}^r h_k(\hat{x}(t)) \Omega_k \hat{x}(t) \quad (20)$$

By combining the uncertain system equations (19), the observer equations (7), the polytopic PDC controller and the estimation errors definitions (8), the following uncertain system with bounded external disturbances is obtained:

$$\dot{x}_a(t) = \sum_{i=1}^r \sum_{j=1}^2 \sum_{k=1}^r h_i(\hat{x}) \mu_j(\hat{a}^u) h_k(\hat{x}) (\Phi_{ijk} x_a(t) + \Psi_{ij} \omega(t)) \quad (21)$$

s.t.  $x_a(t) = \begin{pmatrix} x(t) & e_x(t) & e_{a^u}(t) \end{pmatrix}^T$  represents the augmented (extended) state vector;  $\omega(t) = \begin{pmatrix} a^u(t) & \dot{a}^u(t) \end{pmatrix}^T$  represents the exogenous input (signal attack  $a^u(t)$  and its derivative), supposed unknown but bounded.

Matrices  $\Phi_{ijk}$  and  $\Psi_{ij}$  are defined as follows:

$$\Phi_{ijk} = \begin{pmatrix} \Phi_{ijk}^1 & (\mathcal{B}_{ij} + \Delta B(t)) \Omega_k & 0 \\ \Delta A(t) - \Delta B(t) \Omega_k & A_i - L_{ij} C & 0 \\ 0 & -K_{ij} C & -\alpha_{ij}^u \end{pmatrix} \quad (22)$$

with  $\Phi_{ijk}^1 = A_i - \mathcal{B}_{ij} \Omega_k + \Delta A(t) - \Delta B(t) \Omega_k$  and

$$\Psi_{ij} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \alpha_{ij}^u & I \end{pmatrix} \quad (23)$$

From (21) and the system output equation (1), the resulting closed-loop system becomes:

$$\begin{pmatrix} \dot{x}_a(t) \\ y(t) \end{pmatrix} = \sum_{i=1}^r \sum_{j=1}^2 \sum_{k=1}^r h_i(\hat{x}) \mu_j(\hat{a}^u) h_k(\hat{x}) \begin{pmatrix} \Phi_{ijk} & \Psi_{ij} \\ \bar{C} & 0 \end{pmatrix} \begin{pmatrix} x_a(t) \\ \omega(t) \end{pmatrix} \quad (24)$$

s.t.

$$y(t) = C x(t) = \begin{pmatrix} C & 0 & 0 \end{pmatrix} x_a(t) = \bar{C} x_a(t) \quad (25)$$

Before presenting the stabilization and control conditions, the following definition and lemma are remembered:

**Definition 1** Given a positive scalar  $\gamma$ , the system equation (24) is said to be stable with  $H_\infty$  attenuation level  $\gamma$  if it is exponentially stable with:

$$\int_{\infty}^0 \{(y^T(t))_{\infty}(y(t))_{\infty} - \gamma^2 \omega^T(t)\omega(t)\} dt < 0 \quad (26)$$

where  $\gamma$  is the desired level of disturbance attenuation.

**Lemma 1** Based on Lyapunov theory, the continuous-time system (24) is stable with an  $H_\infty$  disturbance attenuation  $\gamma$  if there exists a positive symmetric matrix  $P = P^T > 0$  s.t.

$$\begin{bmatrix} \Phi_{ijk}^T P + P\Phi_{ijk} & P\Psi_{ij} & \bar{C}^T \\ * & -\gamma^2 I & 0 \\ * & * & -I \end{bmatrix}_{i,k=1,\dots,r,j=1,2} < 0 \quad (27)$$

In order to relax the stability conditions given in Lemma 1, the following formulation is proposed [14], [4]:

**Lemma 2** [15] For a given positive scalar  $\gamma$ , if there exist matrices  $P, Z_{ijk}$ , where  $P = P^T > 0$  and  $Z_{iji}$  are symmetrical,  $Z_{kji} = Z_{ijk}^T$ ,  $i \neq k, i, k = 1, \dots, r, j = 1, 2$  satisfying the following matrix inequalities, then for the uncertain polytopic T-S system (19), the controller (20) makes the  $H_\infty$  norm of fuzzy system (24) less than  $\gamma$

$$\begin{bmatrix} \Phi_{iji}^T P + P\Phi_{iji} & P\Psi_{ij} \\ * & -\gamma^2 I \end{bmatrix}_{i=1,\dots,r,j=1,2} < Z_{iji} \quad (28)$$

$$\begin{bmatrix} (*)^T P + P(\Phi_{ijk} + \Phi_{kji}) & 2P\Psi_{ij} \\ * & 2\Psi_{ij}^T P \end{bmatrix}_{i \neq k, j=1,2} < Z_{ijk} + Z_{kji} \quad (29)$$

$$\begin{bmatrix} Z_{1j1} & \dots & Z_{1jr} & \bar{C}^T \\ \vdots & \ddots & \vdots & \vdots \\ Z_{rj1} & \dots & Z_{rjr} & \bar{C}^T \\ \bar{C} & \dots & \bar{C} & -I \end{bmatrix}_{j=1,2} < 0 \quad (30)$$

By replacing  $\Phi_{ijk}$  and  $\Psi_{ij}$  by their expressions, with some change of variables and classical linearization procedure (Schur's complement and bounded real lemma), the obtained constraints can be easily solved using convex optimization tools and/or the use of a dedicated resolution tool for bilinear constraints like the PenBMI Matlab toolbox (see [16], [17] and [18] for some examples). The proposed solution presents the advantage of a simultaneous design of both the controller and the observer gains using a single-step procedure rather than a classical two-steps procedure of resolution like the one presented in [19].

## 5 Numerical Example

In the following, let us consider the study case of a dynamic modeling and control for quadrotor.

The objective of this work is to ensure the quadrotor safe behaviour and stabilization via an observer based control design. Indeed, stealthy actuator attacks aiming to disturb and destabilize the control and navigation system of the UAV are here considered. These attacks are modeled as unknown but bounded time-varying signals

affecting the system matrix  $B(t)$ .

The first step to this aim will be the quadrotor Polytopic modelling; then, considering the actuator stealthy attacks, the resulting system (attacked one) will be written as an uncertain one, as detailed in previous sections. The proposed control and observer design approach will be then applied to illustrate its efficiency thanks to simulation results.

### 5.1 Polytopic Model of a UAV

In this section, we address the polytopic T-S modelling of a UAV. The considered representation is used in order to rewrite the nonlinear behaviour of the quadrotor into a polytopic-Multiple Model way, without any linearisation, loss of information or approximation. The nonlinear dynamic of the quadrotor is given by the following model:

$$\begin{cases} \ddot{\phi}(t) = \frac{1}{I_x} [(I_y - I_z)\dot{\theta}\dot{\psi} - K_{fax}\dot{\phi}^2 - J_r\dot{\theta}\bar{\Omega} + lU_2] \\ \ddot{\theta}(t) = \frac{1}{I_y} [(I_z - I_x)\dot{\psi}\dot{\phi} - K_{fay}\dot{\theta}^2 - J_r\dot{\phi}\bar{\Omega} + lU_3] \\ \ddot{\psi}(t) = \frac{1}{I_y} [(I_x - I_y)\dot{\theta}\dot{\phi} - K_{faz}\dot{\psi}^2 + lU_4] \end{cases} \quad (31)$$

s.t.  $\bar{\Omega}$  is given by  $\bar{\Omega} = \omega_1 - \omega_2 + \omega_3 - \omega_4$ . The motors control inputs, denoted  $U_i, i = 1, 2, 3, 4$ , are given as a function of the rotors angular velocities as follows:

$$\begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix} = \begin{pmatrix} K_t & K_t & K_t & K_t \\ -K_t & 0 & K_t & 0 \\ 0 & -K_t & 0 & K_t \\ K_d & K_d & K_d & K_d \end{pmatrix} \begin{pmatrix} \omega_1^2 \\ \omega_2^2 \\ \omega_3^2 \\ \omega_4^2 \end{pmatrix} \quad (32)$$

The angles (given in [rad]),  $\phi, \theta$  and  $\psi$  represent the Roll, Pitch, and Yaw angles respectively.

We denote the moment of inertia among axes  $x, y$  and  $z$  as  $I_x, I_y$  and  $I_z$  respectively.

$J_r, K_t$  and  $K_d$  are the rotor inertia, propeller thrust and drag coefficients and  $K_{fax}, K_{fay}, K_{faz}$  the frictions' aerodynamic coefficients. Interested readers can see [1] and [20] for more calculation details. From the SNT transformation, the nonlinear system model (31) can be in a straightforward way written as a quasi-LPV model given by:

$$\begin{cases} \dot{x}(t) = A(t)x(t) + B(t)u(t) \\ y(t) = Cx(t) \end{cases} \quad (33)$$

with suitable state, output and input vectors:

$$x(t) = (\phi \quad \theta \quad \psi \quad \dot{\phi} \quad \dot{\theta} \quad \dot{\psi})^T$$

$$y(t) = (\phi \quad \theta \quad \psi \quad \dot{\phi} \quad \dot{\theta} \quad \dot{\psi})^T$$

$$u(t) = (\omega_1 \quad \omega_2 \quad \omega_3 \quad \omega_4 \quad \omega_1^2 \quad \omega_2^2 \quad \omega_3^2 \quad \omega_4^2)^T$$

the state matrices are given by:

$$A(t) = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\frac{K_{fax}}{I_x}\dot{\phi} & 0 & \frac{I_y - I_z}{I_x}\dot{\theta} \\ 0 & 0 & 0 & 0 & -\frac{K_{fay}}{I_y}\dot{\theta} & \frac{I_z - I_x}{I_y}\dot{\phi} \\ 0 & 0 & 0 & \frac{I_x - I_y}{I_z}\dot{\theta} & 0 & -\frac{K_{faz}}{I_z}\dot{\psi} \end{pmatrix}$$

$$C = I_6$$

and

$$B(t) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{J_x}{I_x} \dot{\theta} & -\frac{J_x}{I_x} \dot{\theta} & \frac{J_x}{I_x} \dot{\theta} & -\frac{J_x}{I_x} \dot{\theta} & -l \frac{K_x}{I_x} & 0 & l \frac{K_x}{I_x} & 0 \\ -\frac{J_x}{I_y} \dot{\phi} & \frac{J_x}{I_y} \dot{\phi} & -\frac{J_x}{I_y} \dot{\phi} & \frac{J_x}{I_y} \dot{\phi} & 0 & -l \frac{K_y}{I_y} & 0 & l \frac{K_y}{I_y} \\ 0 & 0 & 0 & 0 & l \frac{K_d}{I_z} & -l \frac{K_d}{I_z} & l \frac{K_d}{I_z} & -l \frac{K_d}{I_z} \end{pmatrix}$$

Presuming that the variation of angular velocities occurs between known minimum and maximum values, and applying the SNT approach [2], [5], [7], when choosing the following premise variables:

$$z_1(t) = \dot{\phi}, \quad z_2(t) = \dot{\theta}, \quad z_3(t) = \dot{\psi}$$

the resulting polytopic model is then deduced:

$$\begin{cases} \dot{x}(t) = \sum_{i=1}^r h_i(t)(A_i x(t) + B_i u(t)) \\ y(t) = Cx(t) \end{cases} \quad (34)$$

### 5.2 Uncertain System Modelling under Attacks

In the previous subsection, and for the nominal case (no attacks), the polytopic model of the quadrotor-UAV (34) was deduced from its nonlinear dynamics. In the following, the data deception signal will be included and, the global system under attacks will be represented as an uncertain system.

The quasi-LPV system (34) under actuator attacks may be represented as the following:

$$\begin{cases} \dot{x}(t) = \sum_{i=1}^r h_i(t)(A_i x(t) + B_i(t)u(t)) \\ y(t) = Cx(t) \end{cases} \quad (35)$$

where  $B_i(t)$  is given by:

$$B_i(t) := B_i + \Gamma^u d^u(t) \quad (36)$$

Based on the results presented in section 3, the system and attacks observer is given by the system equations (7), and the nonlinear system subject to actuator attacks is modeled thanks to system (19). The objective now is to apply the proposed approach in order to design the robust control law (20) and the observer gains.

### 5.3 Simulation results

The designed observers and controller are implemented and tested through a numerical simulation of a quadrotor robust attitude stabilization despite stealthy actuator data deception attacks.

The design goals and the controller structure (20) based on the state feedback control law is applied to the the nonlinear system equations (33) and (35) subject to the actuator stealthy attacks (36). The control gains are obtained by applying the developed polytopic approach given in Lemma 2 and solving the LMI constraints (28), (29) and (30).

The resulting figures illustrate the stability, robustness and convergence of the system states regarding the attacks. The state, their estimates and estimations error are illustrated in the following figures:

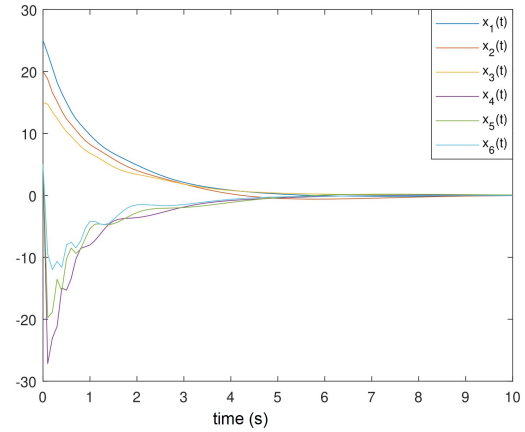


Figure 1: System states estimation errors

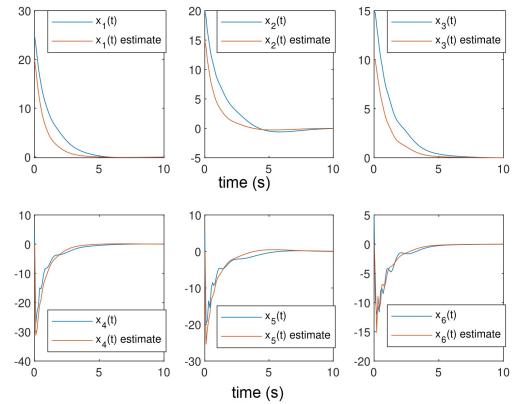


Figure 2: System states estimation errors

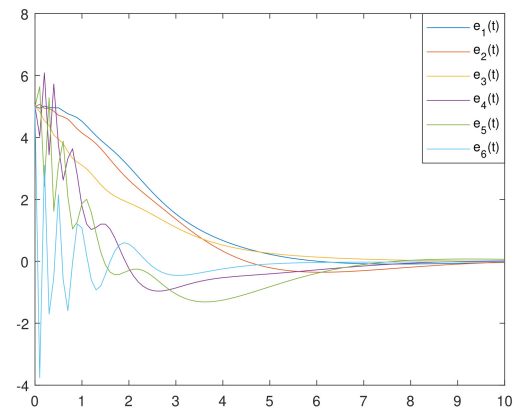


Figure 3: System states estimation errors

The stealthy attack signal  $a^u(t)$  and its estimate is represented in figure 4.

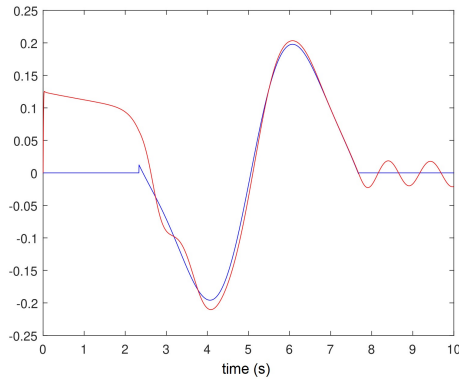


Figure 4: Stealthy attack signal

From the obtained results, one can conclude to the effectiveness of the proposed approach.

Indeed, the system states (angles and velocities) are converging asymptotically despite the stealthy attacks (unknown behaviour), where the estimation errors for both states and data deception attacks tend to zero for the state and with an attenuation level  $\gamma$  for the stealthy signal.

## 6 Conclusion

This paper contribution aimed to propose a polytopic Takagi-Sugeno strategy for the modelling and  $H_\infty$  robust control of a quadrotor subject to stealthy actuator attacks. The attacked UAV system under attacks was modeled as an uncertain polytopic T-S fuzzy one; which allowed us to generalize existing results for the state feedback observer based control design.

The nonlinear system was represented under an uncertain shape (with observable premise variables) allowing the implementation of the observer and control design. The attenuation of the external like perturbation (attack) was guaranteed thanks to the  $H_\infty$  norm. Numerical simulations were given in order to illustrate the effectiveness of the proposed approach. As an extension of this work, a real application example is also under investigation.

**Conflict of Interest** The author declares no conflict of interest.

## References

- [1] S. Bezzaoucha Rebai, "Robust Attitude Stabilization of Quadrotor Subject to Stealthy Actuator Attacks", in the 2022 International Conference on Control, Robotics and Informatics (ICCRI), Danang, Vietnam, April 2-4, 2022, doi: 10.1109/ICCRI55461.2022.00018.
- [2] T. Takagi, M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control", IEEE Transactions on Systems, Man and Cybernetics **15**(1), 116-132, 1985, doi 10.1109/TSMC.1985.6313399.
- [3] M. Sugeno, G. Kang, "Structure identification of fuzzy model", Fuzzy Sets and Systems, **28**(1), 15-33, 1988, [https://doi.org/10.1016/0165-0114\(88\)90113-3](https://doi.org/10.1016/0165-0114(88)90113-3).

- [4] A. Benzaouia, A. El Hajjaji, "Advanced Takagi-Sugeno Fuzzy systems, Delay and saturations", Studies in Systems, Decision and Control 8, 2014, Springer books, <https://doi.org/10.1007/978-3-319-05639-5>.
- [5] K. Tanaka, H. Wang, "Fuzzy Control Systems Design and Analysis: A Linear Matrix Inequality Approach", Ed Hardcover, John Wiley and Sons, Inc., New York, 2001, doi:10.1002/0471224596.
- [6] K. Tanaka, T. Ikeda, H. Wang, "Fuzzy regulators and fuzzy observers: relaxed stability conditions and LMI based designs", IEEE Transactions on Fuzzy Systems **6**(2), 250-265, 1998, doi: 10.1109/91.669023.
- [7] K. Tanaka, T. Hori, H. Wang, "A Multiple Lyapunov Function Approach to Stabilization of Fuzzy Control Systems", IEEE Transactions on Fuzzy Systems **11**(4), 582-589, 2003, doi: 10.1109/TFUZZ.2003.814861.
- [8] T. Guerra, A. Kruszewski, L. Vermeiren, H. Tirmant, "Conditions of output stabilization for nonlinear models in the Takagi-Sugeno's form", Fuzzy Sets and Systems **157**(9), 1248-1259, 2006, <https://doi.org/10.1016/j.fss.2005.12.006>.
- [9] S. Bezzaoucha, H. Voos, M. Darouach, "A contribution to Cyber-Security of Networked Control Systems: an Event-based Control Approach", in 3<sup>rd</sup> International Conference on Event-Based Control, Communication and Signal Processing, Funchal, Madeira, Portugal, 2017, doi: 10.1109/EBCCSP.2017.8022805.
- [10] S. Bezzaoucha Rebai, H. Voos, "Stability Analysis of Power Networks under Cyber-Physical Attacks: an LPV Descriptor Approach", in the 6<sup>th</sup> International Conference on Control, Decision and Information Technologies, Paris, France, 2019, doi: 10.1109/CoDIT.2019.8820425.
- [11] S. Bezzaoucha Rebai, H. Voos, "Simultaneous State and False-Data Injection Attacks Reconstruction for NonLinear Systems: an LPV Approach", in the 3<sup>rd</sup> International Conference on Automation, Control and Robots, ICACR2019, Prague, Czech Republic, 2019, doi: 10.1145/3365265.3365280.
- [12] S. Bezzaoucha Rebai, "A Cyber-Security Contribution to Estimation and Event-Based Control Scheduling Co-Design for Polytopic and T-S Fuzzy Models Using A Lyapunov Approach", Springer Nature - International Journal of Fuzzy Systems, 2022, <https://doi.org/10.1007/s40815-022-01282-3>
- [13] S. Bezzaoucha, B. Marx, D. Maquin, J. Ragot, "Nonlinear joint state and parameter estimation: Application to a wastewater treatment plant", Control Engineering Practice, **21**(10), 1377-1385, 2013, <https://doi.org/10.1016/j.conengprac.2013.06.009>.
- [14] L. Xiaodong, Z. Gingling, "New approaches to  $H_\infty$  controller design based on fuzzy observers for T-S fuzzy systems via LMI", Automatica **39**(9), 1571-1582, 2003, doi: 10.1016/S0005-1098(03)00172-9.
- [15] M. Oudghiri, M. Chadli, A. El Hajjaji, "One step procedure for robust output fuzzy control", in the 15<sup>th</sup> mediterranean conference of control automation, Athens, 1-6, 2007, doi: 10.1109/MED.2007.4433964.
- [16] M. Kocvara, M. Stingl, "PENNON – a code for convex nonlinear and semidefinite programming", Opt. Methods and Software, **18**(3), 3170-333, 2003, <https://doi.org/10.1080/1055678031000098773>.
- [17] M. Kocvara, M. Stingl, "PENBMI, Version 2.0", See [www.penopt.com](http://www.penopt.com) for a free developer version, 2004.
- [18] D. Henrion, J. Lofberg, M. Kocvara, M. Stingl, "Solving polynomial static output feedback problems with PENBMI", Proceedings of the 44th IEEE Conference on Decision and Control, 7581-7586, 2005, doi: 10.1109/CDC.2005.1583385.
- [19] J. Lo, M. Lin, "Observer-based robust  $H_\infty$  control for fuzzy systems using two-steps procedure", IEEE Trans Fuzzy Systems, **12**(3), 350-359, 2004, doi: 10.1109/TFUZZ.2004.825992.
- [20] F. Torres, A. Rahbi, D. Lara, G. Romero, C. Pégard, "Fuzzy State Feedback for Attitude Stabilization of Quadrotor", International Journal of Advanced Robotic Systems, InTech, 2016, <https://doi.org/10.5772/61934>.

## An Ensemble of Voting- based Deep Learning Models with Regularization Functions for Sleep Stage Classification

Sathyabama Kaliyapillai<sup>\*1</sup>, Saruladha Krishnamurthy<sup>1</sup>, Thiagarajan Murugasamy<sup>2</sup>

<sup>1</sup>Computer Science and Engineering, Pondicherry Engineering College, Pondicherry, 605014, India

<sup>2</sup>Neyveli Lignite Corporation Ltd, Neyveli, India

### ARTICLE INFO

Article history:

Received: 14 November, 2022

Accepted: 07 January, 2023

Online: 30 January, 2023

Keywords:

Sleep stage classification

Activity Regularization

LSTM

GRU

RNN

Ensemble Voting method

DL model

### ABSTRACT

Sleep stage performs a vital role in people's daily lives in the detection of sleep-related diseases. Conventional automated sleep stage classifier models are not efficient due to the complexity linked to the design of mathematical models and extraction of hand-engineering features. Further, quick oscillations amongst sleep stages frequently lead to indistinct feature extraction, which might result in the imprecise classification of sleep stages. To resolve these issues, deep learning (DL) models are applied, which make use of many layers of linear and nonlinear processing components for learning the hierarchical representation or feature from input data and have been used for sleep stage classification (SSC). Therefore, this paper proposes an ensemble of voting-based DL models, namely the recurrent neural network (RNN), long short-term memory (LSTM) and gated recurrent unit (GRU), with activation Regularization (AR) functions for SSC. The penalty addition of L1, L1\_L2, and L2 on the layers of the model fine-tunes it in proportion to the magnitude of the activation function in the model by reducing overfitting. Subsequently, the presented model integrates the results of every classification model to the max voting combination rule. Finally, experimental results of the proposed approach using the benchmark Sleep Stage dataset are evaluated using various metrics. The experimental results illustrates that the Ensemble RNN, Ensemble GRU, and Ensemble LSTM models have achieved an accuracy of 85.57%, 87.41%, and 89.01%, respectively.

## 1. Introduction

Sleep acts as a vital part of the physical health and quality of human lives. Sleep diseases, like obstructive sleep apnea (OSA) and insomnia, might result in daylight drowsiness, depression, or even mortality [1]. Thus, there is a need to design an efficient method for diagnosing and treating sleep-related diseases. The study of sleep-related diseases is labeled "sleep medicine," which was once a significant medical field and has been included in various medical challenges. It consists of two major categories of sleep namely, rapid eye movement (REM) sleep and nonrapid eye movement (NREM). The REM and NREM sleep stages are adjacent and alternated by the sleep procedure on a regular basis, and unbalanced cycling or the absence of a sleep stage results in a sleep disorder [2]. Inappropriate sleep disorders result in inferior quality of sleep, which is frequently ignored and emphasized that sleep-related issues are a forthcoming worldwide health problem

[3]. Sleep stage classification is the initial phase of the diagnosis of sleep-related diseases[4]. The critical stage in sleep research is collecting the polysomnographic (PSG) information from the patient at the time of sleep. The PSG information comprises electromyography (EMG), electroencephalography (EEG), biophysiological signal, electrocardiography (ECG), and respiratory efforts. Until recently, the sleep stage score had to be physically determined by human experts [5].

A human expert's ability to manage slower changes in background EEG is limited, and he or she learns the distinct guidelines for scoring sleep stages from multiple PSG recordings [6]. Moreover, the calculations by the sleep expert are inclined to inter and intra- observer variability, which influences the quality of the sleep stage score. This substantiates the need for sleep stage scoring using Artificial Intelligence (AI) techniques[7].

Sleep stage classification has been studied for several years, and different advanced techniques and medical application areas

\* Corresponding Author: Sathyabama Kaliyapillai, sathii\_manju@pec.edu

have been established. ML techniques used for SSC are artificial neural networks (ANN), support vector machines (SVM), dual-tree, K-means clustering, and empirical mode decomposition (EMD). However, these traditional methods rely on the detection of biological signals [8]. The manual features are created from the EEG signal, which has a tendency toward local optimization. Moreover, the patterns of brain signals are complex compared to the present knowledge of human beings, which might result in data loss in the manual way of extracting features. Additionally, feature extraction is a difficult and lengthy process. It also necessitates excessively long working hours for experienced experts. The convenience and accuracy of sleep stage classification techniques are critical issues in the analysis of sleep-related diseases.

In recent times, different studies have utilised deep learning (DL) models, which are motivated by the biological imitation outcomes of the visual cortex of mammals. In contrast to the conventional technique, it decreases the difficulty of the network and weight count due to its shared weight networking model, which is equivalent to a biological NN. Additionally, it reduces the calculation process because of its capability of classifying the EEG data without hand-crafted feature extraction.

This paper presents DL models with AR regularisation functions and ensemble DL models for sleep stage classification. At the initial stage, the required features were extracted from the single channel and normalized. Following this, the proposed model employed three DL models, namely, the recurrent neural network (RNN), long short-term memory (LSTM), and gated recurrent units (GRU), for classifying the sleep stages. At last, the presented model integrates the results of every classification model using the max voting combination rule to generate an optimal outcome. The experimental analysis was performed to highlight the improvements of the proposed model over the existing models.

The construct of the paper is detailed as follows: Section 2 summarizes an overview of the existing work based on deep learning techniques for sleep stage classification. Section 3 provides an overview of the proposed work for SSC using DL models, various Regularization, and ensemble techniques. Section 4 discusses the dataset details, implementation details, and performance evaluation of the proposed work. Section 5 provides conclusion on performance on proposed model on sleep stage dataset.

## **2. Related works**

The author proposed an NN-based CNN with an attention scheme for automated sleep stage classification. The weighted loss function employed in the CNN model handled the class imbalance problem for sleep stage classification [9]. Developed an automated DL-based sleep stage classification model utilizing EEG signals that automatically extracted the time-frequency spectra of the EEG signals. The Continuous Wavelet Transform (CWT) technique was used for extracting the RGB color images of the EEG signal. The transfer learning of the pre-trained CNN was utilised to classify the CWT images according to sleep levels [10].

Developed an orthogonal convolutional neural network (OCNN) for learning rich and efficient feature representation. The Hilbert-Huang transform was used to extract the time-frequency representation of the EEG signal, and OCNN was used to classify the sleep stages [11]. An effective ensemble method to classify distinct types of sleep stages. The classification technique was employed an integration of the EEGNet and BiLSTM models for learning the distinct features of EEG and EOG signals, respectively [12].

In the past few decades, the sleep stage classification process has gained significant attention. Machine learning techniques such as multiclass SVM, and linear discriminant analysis were applied for classifying sleep stages [13]. Proposed a technique for detecting sleep stages based on iteration filtering. The amplitude envelope and instantaneous frequency (AM-FM) were applied for classifying the sleep stages, and an average accuracy of 86.2% across five sleep stages was achieved[14].

The author proposed a novel sleep stage recognition method based on a new set of wavelet-based features extracted from massive EEG datasets. The integrated SVM technique and CNN model were employed on the EEG signal for extracting features and classifying the sleep stages. It was implemented to learn task-oriented filters to classify data depending on single-channel EEG without utilizing previous domain information [15].

The author proposed a deep CNN framework extracted data from raw EEG signals and classified the sleep stages using the SoftMax activation function[16]. Smart technology for sleep stage classification was developed, data were trained using two different fuzzy rule algorithms for classifying sleep stages and studying the new patient's record. But it ignores the connection between the current stage and its adjacent sleep stage and does not capture the transition rules among the sleep stages [17].

An Elman RNN was proposed for automatically classify sleep staging systems. This system classified different sleep stages based on energy features (E) of 30 s epochs extracted from a single channel's EEG signals [18]. The author proposed DeepSleepNet model extracted time-invariant features from the EEG signal using CNN and bi-directional LSTM and learned the stage transitions rule. Also, the two-step training algorithm was used to lessen class-imbalance problems and encode the temporal information of the EEG signal into the model [19]. The author developed a mixed neural network (MLP and LSTM), the temporal physiological features of the signal were extracted using power spectral density (PSD), and the extracted features were classified using an MLP and LSTM [20]. The sequential feature learning model was developed using a deep bi-directional RNN with an attention method for single-channel automated sleep stage classification. The time-frequency features were extracted from the EEG [21].

The sleep stage classifier technique was proposed, the temporal (59) and frequency domain (51) characteristics of the EEG signal were extracted using the PSD approach, and the extracted features were classified using the C-CNN and attention-based BiLSTM models [22]. The author proposed SleepEEGNet combines the CNN and BiLSTM models to extract the time and frequency features and capture the sleep transition between the epochs in a single-channel EEG signal. The new loss function

technique in the SleepEEGNet decreases the effects of the class imbalance problems [23].

The author developed a classification framework for automatic sleep stage recognition from a combination of male and female human subjects. Then the ResNet structure automatically extracts the frequency features from the raw EEG signal [24].

Transfer learning-CNN was developed for classifying the sleep stages. The time and frequency features were calculated using PSD estimation and statistical techniques from the EEG and EOG signals. The EEG feature set and the set of fused features of EEG and EOG signals were separated and converted into image sets using a horizontal visibility graph (HVG) in Euclidean space. An image of HVG is classified into different sleep stages using transfer learning-CNN [25].

### 3. Proposed Sleep Stage Classification Model

The workflow involved in the proposed model for SSC is given in Figure. 1. The figure shows that the initial stages of the processing of input EEG data involve data extraction of sub-band frequency and data normalization. Followed by three DL models with activity regularization techniques are used for the classification of EEG signals for SSC. Finally, the max voting ensemble method is applied to determine the performance of the optimal sleep state classification results of the presented model.

#### 3.1. Data acquisition and Preprocessing

The multichannel time series data is extracted from different channels of EEG (Fpz-Cz), Pz-Oz, and EOG. The EEG signal is recorded by positioning the electrode in accordance with the International 10-20 systems. The EEG data from a single EEG channel (Fpz-Cz) is considered for this research work. The steps for extracting the EEG signal data that is fed into the DL model are narrated as follows:

- The extracted EEG signal (time series) of 30-sec epochs is fed as input to the DL models.
- The continuous raw signal is converted into sequential data of 30 s epochs is segmented, and stages of S1, S2, S3/4, wake, and REM are assigned in each epoch based on the annotation file in the AASM standard.
- Since each segment(fragment) of 30 s epochs was sampled at 100Hz, and 3000- time points (30\*100) for five stages, are extracted.

The power spectral density (PSD) technique is applied to extract different sub-bands frequencies (35 features) from the EEG signal to identify each stages correctly. The signal is then normalized to have a zero mean and unit variance for each of the 30-second epochs and divided by each segment's power spectral density of each frequency band (0.5 to 30 Hz) by each segment's total power spectral density. The power spectral intensity of the kth is measured by Eq. (1).

$$PSI_k = \sum_{i=\lfloor N(f_k/f_s) \rfloor}^{\lfloor N(f_{k+1}/f_s) \rfloor} |X_i|, \quad k = 1, 2, \dots, K - 1 \quad (1)$$

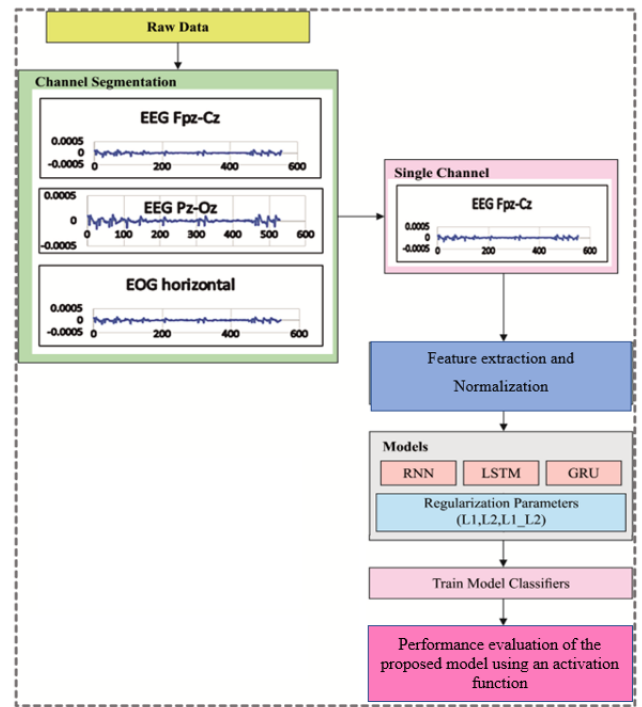


Figure 1: Overall Process of Proposed DL model

#### 3.2. DL Models

The DL models LSTM, GRU, and RNN are discussed in the following section.

##### 3.2.1. RNN Model

RNN is a kind of NNs with loops that permits persisting data from the past in the neural network system. In Figure. 2, the center square signifies a NN that takes input  $x_t$  at present time step  $t$  and provides the value  $h_t$  as an outcome. The loop in the model allow to utilize data from the previous time step for producing output at the present time like step  $t$ . So, it is the state that the decision develops at time slice  $t - 1$  influences the decisions to be developed at time step  $t$ . Thus, the RNN output for the novel information is based on the present input and recent past output data [26]. The RNN output computation depends on the frequent computation of the outcome using Eqn. (2)-(3):

$$h_t = H(W_t x_t + W_h h_{t-1} + b_h) + AR \quad (2)$$

$$y_t = W_y h_t + b_y \quad (3)$$

where  $x_t$  implies the input order at the current time step  $t$ ,  $y_t$  represents the output order at time step  $t$ , and  $h$  signifies the order of the hidden vectors in the time step 1 to T.  $W$  and  $b$  denotes weight matrix as well as bias correspondingly.

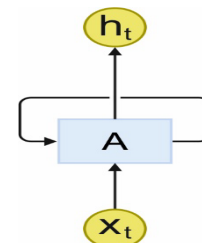


Figure 2: Loop structure of RNN



### 3.2.2. LSTM Model

Hochreiter and Schmidhuber introduced the LSTM networks in 1997 [27]. LSTM is a different kind of RNN with memory cells. These memory cells are important to manage long-term dependencies on the information. The chain like architecture of LSTM is as shown in Figure 3 and specific memory cells from LSTM. All large square blocks imply the memory cell. The cell states are an important portion of the LSTM model which are represented through the horizontal line moving with the top of cell from the Figure. It executes in all cells from the chain of LSTM networks. The LSTM takes the possibility of adding or deleting data in the cell state. This function is completed by other architecture in LSTM known as gates. The gates are computed using the sigmoid  $\sigma$  activation function (demonstrated by  $\sigma$  in Figure. 3) and point-wise multiplication function (illustrated  $\otimes$  in Figure. 3). They are 3 gates that manage data to pass with the cell state. The forget gate is responsible to remove information from the cell state. Besides, the input gate is accountable for appending information to the cell state. The output gate elects the data of the cell state to the outcome.

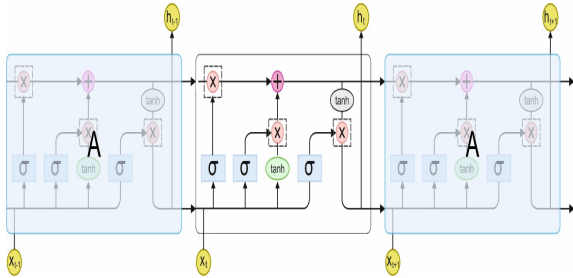


Figure 3: Architecture of LSTM model

The computation in the typical single LSTM cell can be expressed by:

$$\begin{aligned}
 f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) & (4) \\
 i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) & (5) \\
 \tilde{C}_t &= \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) & (6) \\
 C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t & (7) \\
 o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) + AR & (8) \\
 h_t &= o_t * \tanh(C_t) & (9)
 \end{aligned}$$

where the activation function utilized is sigmoid function ( $\sigma$ ) and hyperbolic tangent function ( $\tanh$ ),  $i_t, f_t, o_t$  signifies the input gate, forget gate, output gate respectively,  $C_t, \tilde{C}_t, h_t$  memory cell current content, new cell state, hidden state correspondingly. Every  $W$  and  $b$  refer to the weight matrix and bias, respectively.

### 3.2.2. GRU Model

The GRUs are the other kind of RNNs with memory cells. The GRU also takes a gating scheme for controlling the flow of data with cell state but takes few parameters and does not comprise an output gate. The GRU has 2 gates,  $r$  implies the reset gate, and  $z$  represents the update gate as is shown in Figure 4. The reset  $r$  gate controls the new input data and decides how much of the past information should be forgotten. The update gate updates the information of the previous state and carries that information (data) for a prolonged period [28].

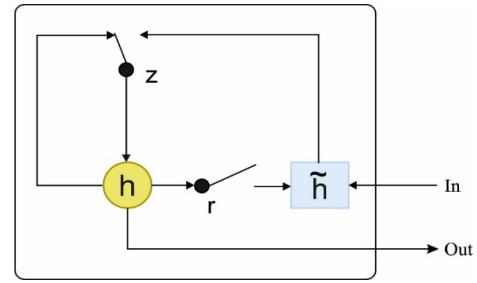


Figure 4: Structure of GRU model

The subsequent formulas are utilized in GRU outcome computations:

$$\begin{aligned}
 r_t &= \text{sigm}(W_x x_t + W_h h_{t-1} + b_r) & (10) \\
 z_t &= \text{sigm}(W_x x_t + W_h h_{t-1} + b_z) + AR & (11) \\
 \tilde{h}_t &= \tanh(W_x x_t + W_h (r_t \odot b_z) + b_h) & (12) \\
 h_t &= z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t & (13)
 \end{aligned}$$

In Eqs. (10)-(13),  $x_t, h_t, r_t, z_t$  implies the input vector, output vector, reset gate, and an updated gate correspondingly. Every  $W$  variable refers to the weight matrix, and  $b$  signifies the bias. The following section discusses various regularization techniques.

### 3.3. Regularization Functions

Overfitting is a prominent issue in the deep learning model, which prevents from completely generalizing the models to fit perfectly on the validation set during training. During the initial stage of training, the validation error decreases typically along with the error on the training set. However, the validation set error will increase as the model starts to overfit the data. Overfitting in the learning curve while training the model is as shown in Figure 5. The learning curve is a graphical plot of learning the data and diagnosing the model's learning performance through loss values (y-axis) with respect to epochs (x-axis). The performance of the deep learning model creates a vast gap, resulting in random fluctuations between the training loss (high performance) and the validation loss (lower performance) while training and evaluating the model.

The overfitting of data happens because of the following reasons.

The model comprises of more than one hidden layer stacked together with nonlinear information processing to learn the association between input and output data and the learning of the association is a complex process.

- Additionally, deep neural networks' loss function/cost function is highly nonlinear and not convex [29].

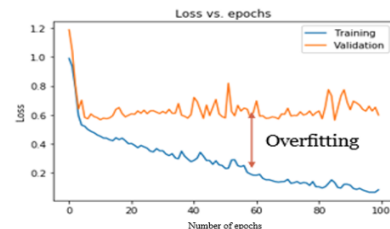


Figure 5: Overfitting (Learning curve)

### 3.3.1. Purpose of regularization

In the literature, to overcome the overfitting problem, various regularization techniques are adopted. The “activity regularization” technique is applied to the DL models to improve the performance to a great extent, mainly when an overfitting problem occurs [30]. It can be applied either to the hidden layers or the output layers of the DL models. It aids minor changes in the weight matrix of the learning approach while learning the data and thus reduces generalisation errors.

### 3.4. Various Regularization techniques

In this section various activation regularization techniques are discussed below.

**L1 activation (Activity) regularization (AR):** The L1(AR) technique is applied to the activation function in the DL model. L1 Regularization is calculated as the "sum of the absolute activation values." The L1 AR causes the activation values to be sparse, thus allowing specific activations to reach zero. The L1 norm may be a more commonly utilised activation Regularization penalty [31].

**L2 activation (Activity) regularization:** L2 (AR) Regularization is calculated as the "sum of the squared activation values." L2 Regularization keeps the magnitude of activations small, allowing specific activation values other than zero [32].

In this research work, L1, L2, and L1\_L2 Regularization techniques alone are used for the experiments, which aid in better decision-making and prediction. This technique aids in improving the learning process in the DL models, thereby reducing generalisation errors.

### 3.4. Ensemble techniques

The ensemble technique combines the decisions/predictions from multiple models to make a final prediction and is used to enhance the model's performance. The simple ensemble techniques of majority voting is as shown in Figure 6.

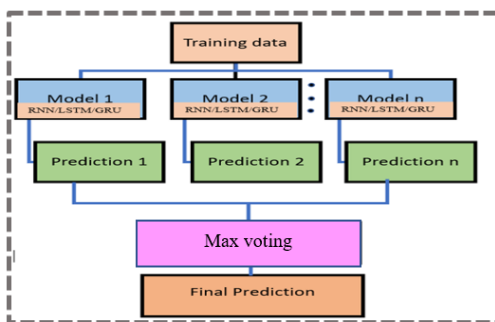


Figure.6: Simple Ensemble Techniques

### Majority (max) ensemble voting

In max voting technique, the output of the multiple DL models is combined using the max-voting technique to make final predictions/decisions. The model classifies the instance to 1 and 0 otherwise for the  $j^{th}$  class of the  $t^{th}$  model [33].

$$\sum_{t=1}^M d_{t,k} = \max_{j=1}^C \sum_{t=1}^M d_{t,k} \quad (14)$$

where  $t = 1, 2, \dots, M$ ,  $M$ - is the number of model classifiers and  $k=1, 2, \dots, C$ ,  $C$ - is the number of classes.

## 4. Implementation

### 4.1. Dataset Details

The sleep staging datasets from the physionet consist of 197 recordings of PSG signals, including 153 sleep cassettes (SC) of healthy patients and 44 sleep telemetry (ST) patients with medication. The details associated with the sleep dataset for 15 subjects are as given in Table 1. The sleep dataset contains bipolar channels (Fpz-Cz and Pz-Oz). The single channel (Fpz-Cz) indicates that the brain activity related to sleep stage connectivity is located in the cerebral midline. The DL model quickly learns the sequential features from a single channel (Fpz-Cz) to minimize the processor’s load and computational time. The channel selection process involves choosing a single channel for the sleep stage classification process. This work using three DL models to automatically classify sleep stages using a single channel (Fpz-Cz) from EEG signals (physionet.org).

Table 1. Dataset Details

Dataset	Wake (W)	S1 (N1)	S2 (N2)	S4 (N3-N4)	REM	Total Epochs
Sleep-EDF-18	8006	635	3621	1299	1609	15,170

In this dataset, 10% of patients do not have alpha waves during w. Sleep stage scoring is a time series (sequential) problem, so it depends on temporal features and previous epochs of the sleep stages (the N2 stage depends on the N1). The benchmark sleep stage dataset (physionet) was used in the experiment to assess the performance of the DL models. This research work used recordings of data from fifteen (15) subjects, ages 25 to 101. The original recording consists of sleep stages labeled as W (wake), 1, 2, 3, 4, M (movement time), R (REM), and unknown (?). For experimental purposes, only five stages, viz., wake, REM, 1, 2, 3, and 4, are considered. In addition, movement time and unknown stages are not taken into consideration. Stages 3 and 4 are considered a single stage according to AASM standards. The DL model's performance is measured using accuracy, recall, F-score, precision, and kappa coefficient.

### 4.2. Platform used for Implementation.

Keras is one of the deep learning libraries that supports the implementation of complicated pre-packaged architectures like RNN, GRU, and LSTM. The DL model experiments were conducted on an Intel Core i5 processor with 16 GB of RAM. The deep learning models were developed using the Python programming language. The training parameters for the SSC dataset are tabulated in Table 2. The parameters of each DL model

were fixed by conducting several experiments and considering various combinations; the model that produced the best results was saved for this research work.

Table 2. Experimental design and Training parameters

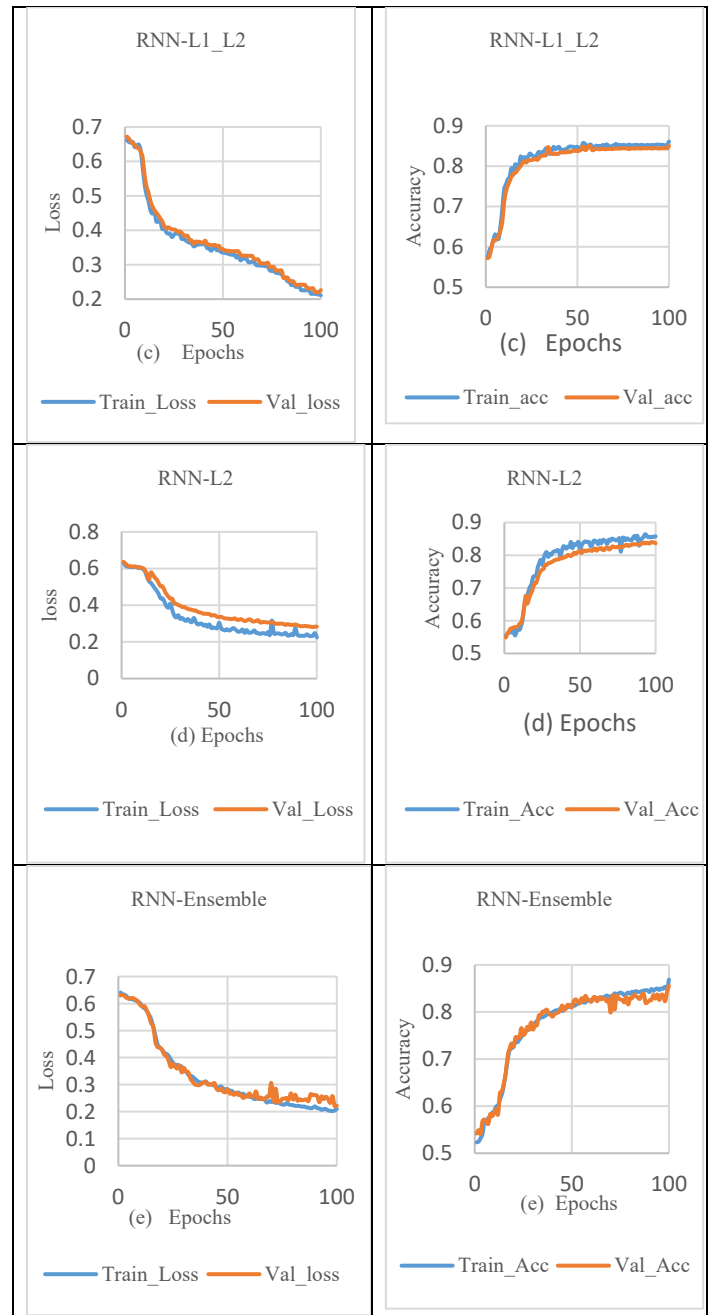
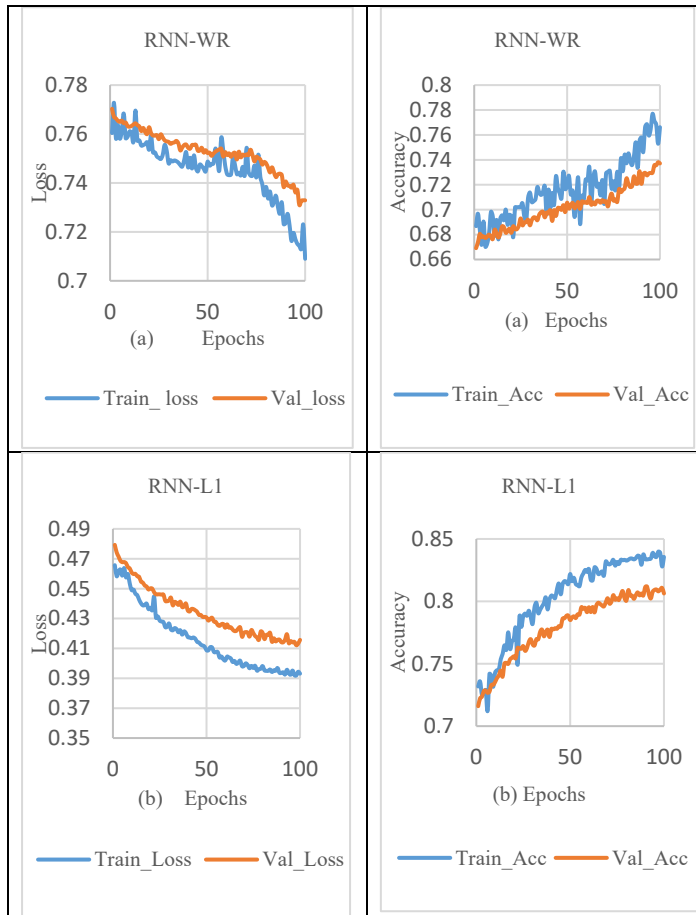
Parameter	Value
Batch size, Epochs, and optimizer	32, 100 and SGD respectively
Layer one	Sequential input layer
Layer 2 RNN/LSTM /GRU	90 (Number of neurons)
Layer 3 RNN/LSTM /GRU	50 (Number of neurons)
Layer 4 Fully connected layer	10 (Number of neurons)
Layer 5 Output layer	5 SoftMax AF

### 5. Performance Evaluation

#### 5.1. Experiments using RNN with (or) without regularization (WR)

The comparative result analysis of RNN model is evaluated with and without regularization as depicted in Table 3. The performance of the model is computed in terms of precision-recall, f-measure, training loss, validation loss, validation accuracy and training accuracy is given in Table 7. From the graph shown in Table 3, the performance of RNN-WR (without regularization) shows that there is a high gap and random fluctuation between validation loss and training loss, which indicates the onset of overfitting, as shown in Tables 3 (a) and 7.

Table 3. RNN learning curve with and without regularization



In order to overcome the overfitting problem in the RNN model, the L1 norm activity regularisation technique with a penalty value of 0.001 was applied to the RNN layer. It is observed that validation loss is reduced but fails to close the gap between the training loss and validation loss in the sleep stage classification process, as shown in Tables 3 (b) and 7. In addition, when the L1\_L2 norm activity regularization technique with a penalty value of 0.001 was set to the RNN layer, the significant gap between the training and validation loss were minimised, which increased the validation accuracy, as shown in Table 3. (c) and 7.

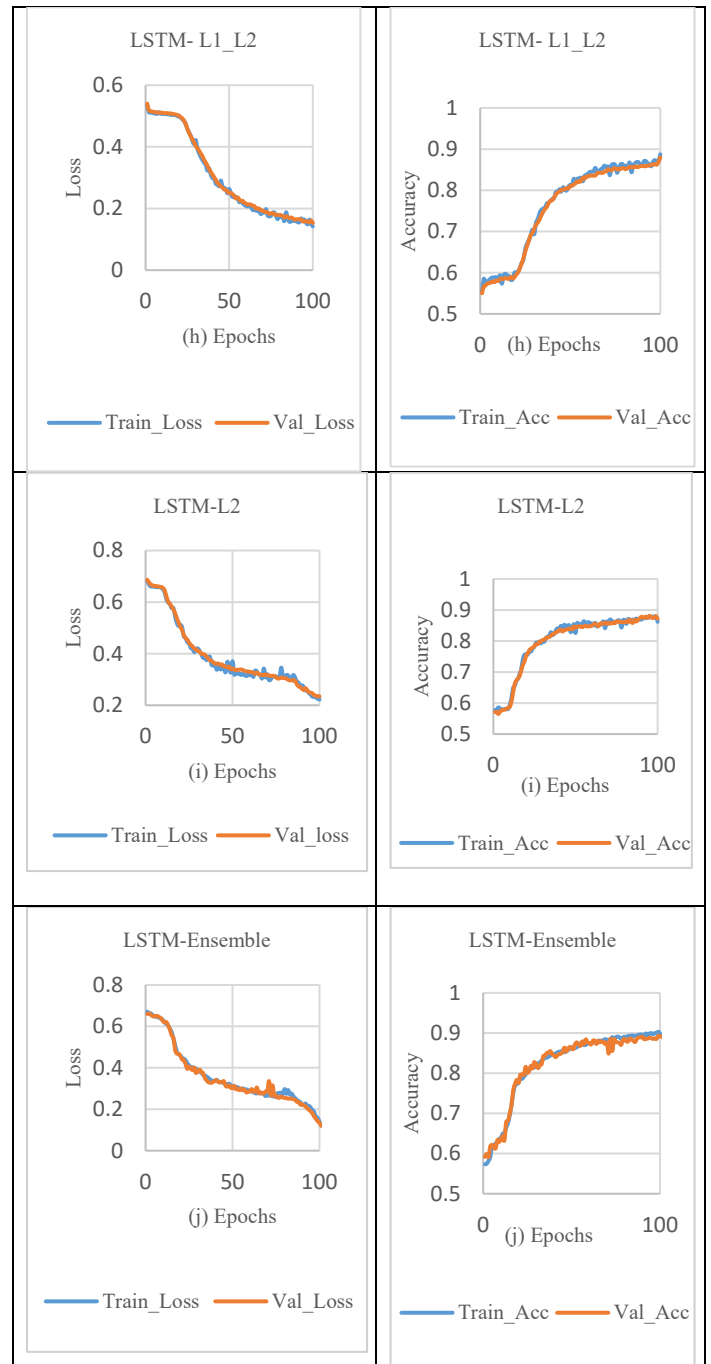
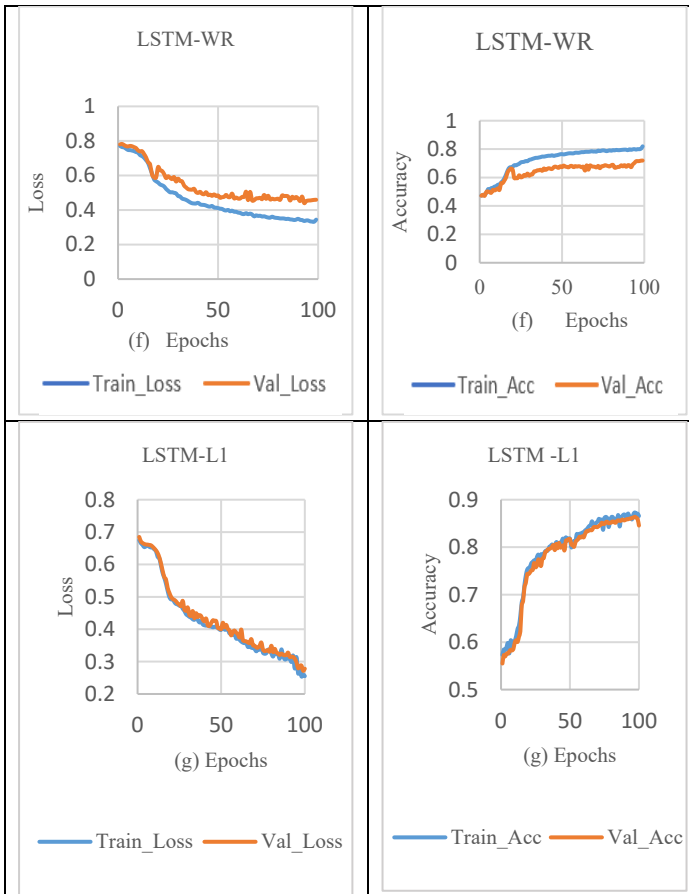
Besides, the L2 norm activity regularization technique with a penalty value of 0.001 to the RNN layer, reducing validation loss, and slightly closing the gap between the training loss and validation loss. So, the RNN model with L2 regularization achieved lower validation loss and higher validation accuracy, as

shown in Table 3. (d) and 7. The RNN models with and without the regularization effect were integrated using the max voting technique to make the final prediction of ensemble models. At last, ensemble RNN achieved lower validation and training loss indicates no sign of overfitting as shown in Table 3 (e). Also, the ensemble model exhibit on-par performance with the effect of adding combined regularization (L1\_L2) for sleep stage classification in the RNN model, as shown in Table 3. (c) and 7.

5.2. Experiments using LSTM with (or) without regularization

Table 4 shows the LSTM model's comparative result analysis. The LSTM model results are evaluated using metrics such as training loss, validation accuracy, validation loss and training accuracy, which are also computed and reported. As shown in Tables 4 (f) and 7, the performance of LSTM-WR (without regularization) for sleep stage classification during training predicts the output with a lower training loss and a higher validation loss, indicating the sign of overfitting. Overcome the overfitting problem in the LSTM model, the L1 norm activity Regularization technique with a penalty value of 0.001 was applied to the LSTM layer. It is observed from Table 7 that LSTM with L1 Regularization achieved a loss difference of 0.0231, which indicates validation loss is reduced and slightly closes the gap between the training and validation losses in the sleep stage classification, as shown in Tables 4 (g) and 7.

Table 4. LSTM learning curve with and without regularization



In addition, the L1\_L2 norm activity Regularization technique with a penalty value of 0.001 was applied to the LSTM layer. It is observed in Table 7. that LSTM with L1\_L2 activity Regularization achieved a loss difference of 0.0111, which effectively closed the significant gap between the training and validation losses, thus increasing validation accuracy, as shown in Tables 4(h) and 7.

Besides, the L2 norm activity Regularization technique with a penalty value of 0.001 was applied to the LSTM layer. It is observed from Table 7 that LSTM with L1 Regularization achieved a loss difference of 0.012, reducing validation loss and closing the gap between the training loss and validation loss. So, the LSTM model with L2 regularization achieved lower validation loss and higher validation accuracy, as shown in Tables 4 (i) and 7.

The LSTM models with and without the Regularization effect are combined using the max voting technique to make the final prediction of ensemble models. At last, ensemble LSTM attained higher performance with lower validation and training losses, which achieved a loss difference of 0.0013, indicating no sign of overfitting, as shown in Tables 4 (j) and 7. To conclude, the ensemble LSTM model exhibits higher performance and closes the gap between training loss and validation loss for SSC.

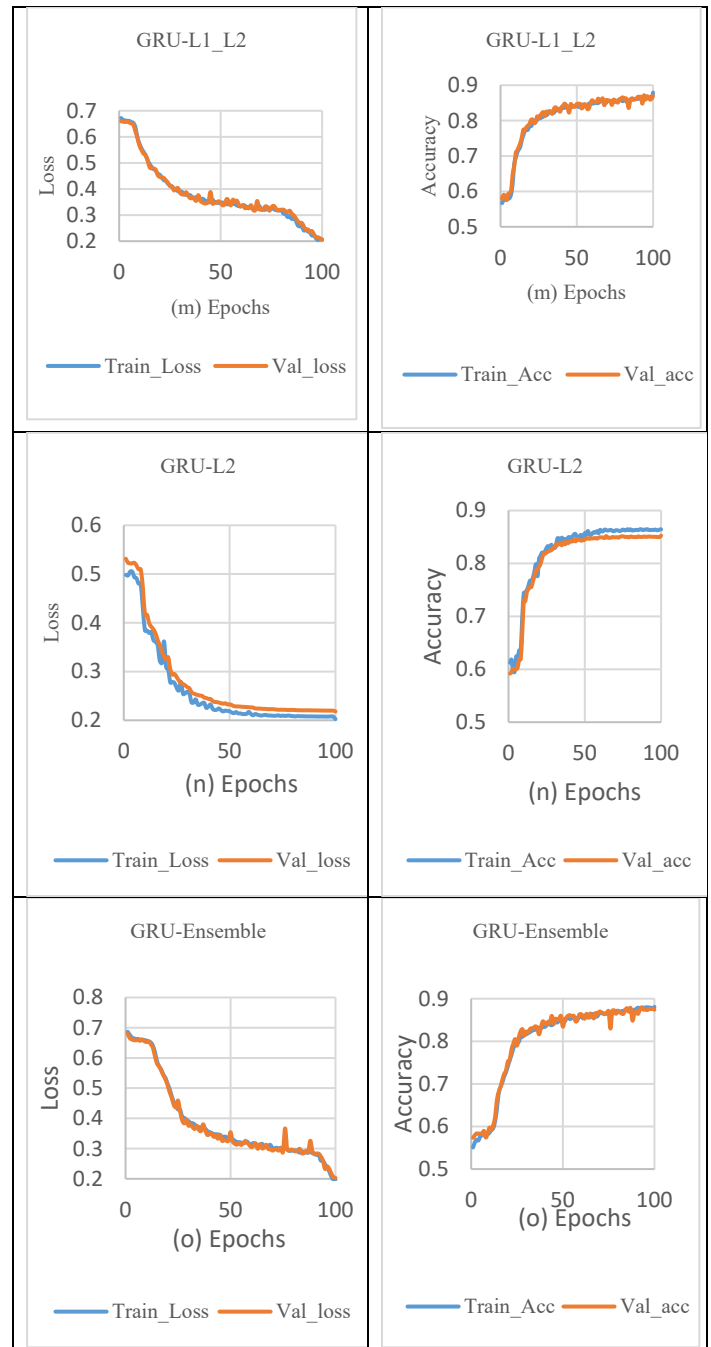
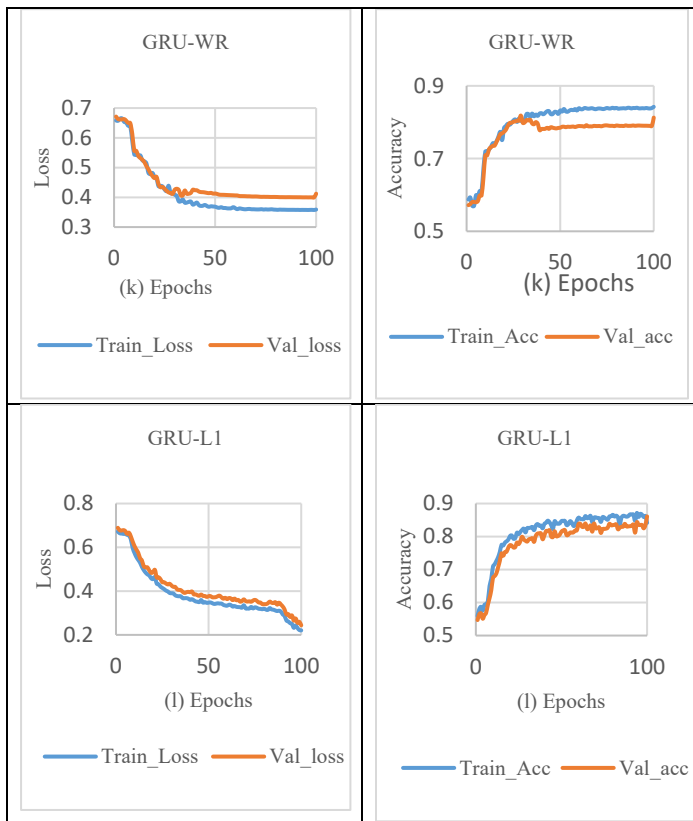
5.3. Experiments using GRU with (or) without regularization

Table 5 depicts the GRU model's comparative result analysis. From the graph shown in Table 5, the performance of GRU-WR (without Regularization) for the sleep stage classification model attained a lower training loss and a higher validation loss, which discloses the sign of overfitting, as shown in Tables 5 (k) and 7. Overcome the overfitting problem in the GRU model, the L1 norm activity Regularization technique with a penalty value of 0.001 was applied to the GRU layer.

It is observed from Table 7 that GRU with L1 Regularization achieved a loss difference of 0.0231, which indicates validation loss is reduced but failed to close the gap between training and validation loss in the sleep stage classification process, as shown in Tables 5 (l) and 7.

In addition, the L1\_L2 norm activity regularization technique with a penalty value of 0.001 was applied to the GRU layer. It is observed from Table 7 that GRU with L1\_L2 Regularization achieved a loss difference of 0.0018, which effectively closed the significant gap between the training and validation loss, thus increasing validation accuracy, as shown in Tables 5 (m) and 7.

Table 5. GRU learning curve with and without Regularization



Besides, the L2 norm activity Regularization technique with a penalty value of 0.001 was applied to the GRU layer. It is observed from Table 7 that GRU with L1 Regularization achieved a loss difference of 0.0158, reduced validation loss, and slightly closed the gap between the training loss and validation loss. So, the GRU model with L2 Regularization achieved lower validation loss and higher validation accuracy, as shown in Tables 5. (n) and 7.

The GRU models with and without the regularization effect are combined using the max voting technique to make the final prediction of ensemble models. At last, ensemble GRU attained higher performance with lower validation and training loss, which achieved a loss difference of 0.0052, indicating no sign of overfitting, as shown in Tables 5 (o) and 7.

Table 7. Result analysis of DL models on sleep stage dataset

Models	Precision (%)	Recall (%)	F-Measure (%)	Training Loss	Training Accuracy (%)	Validation Loss	Validation Accuracy (%)
RNN-WR	74.33	76.67	75.48	0.7090	76.62	0.7329	73.71
RNN-L1	83.12	84.52	83.81	0.3310	83.62	0.4730	80.92
RNN-L1L2	85.41	89.21	87.27	0.2112	86.10	0.2264	85.03
RNN-L2	87.01	87.13	87.06	0.2242	85.78	0.2831	83.90
Ensemble RNN	85.82	89.84	87.78	0.2093	86.88	0.2221	85.57
GRU-WR	82.71	86.42	84.52	0.3115	84.42	0.4123	81.27
GRU-L1	88.09	89.36	88.72	0.2205	86.10	0.2436	84.14
GRU-L1L2	87.84	90.09	88.95	0.2060	87.90	0.2078	86.95
GRU-L2	87.94	89.48	88.7	0.2023	86.45	0.2181	85.25
Ensemble GRU	88.45	89.88	89.16	0.1977	88.08	0.2029	87.41
LSTM-WR	85.48	87.10	86.28	0.3533	85.45	0.5682	81.97
LSTM-L1	87.98	88.78	88.38	0.2551	86.56	0.2782	84.55
LSTM-L1L2	88.10	89.07	89.38	0.1420	88.75	0.1531	87.98
LSTM-L2	89.12	87.34	88.22	0.2225	87.21	0.2345	86.17
Ensemble LSTM (E-LSTM)	88.98	90.76	89.86	0.1201	89.18	0.1214	89.01

Table 7 indicates the sleeping stage classification outcome of the different DL models with ensemble techniques. Figures. 7 illustrates the result analysis of different DL models with ensemble techniques on the sleep stage dataset.

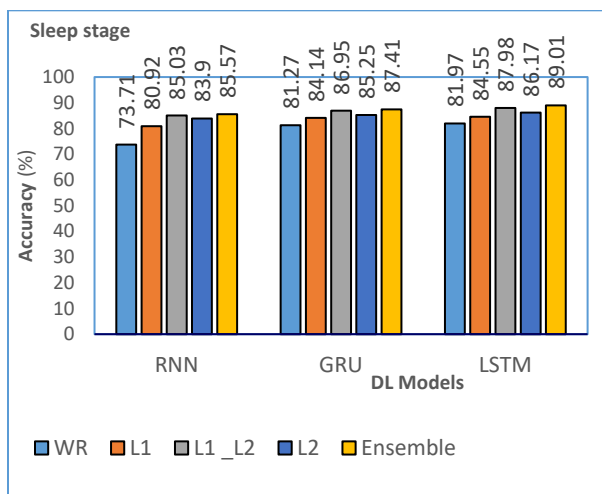


Figure. 7: Accuracy analysis of DL models for sleep stage

The ensemble models such as Ensemble RNN, Ensemble GRU, and Ensemble LSTM models have accomplished maximum validation accuracy of 85.57%, 87.41%, and 89.01%, respectively. Among the three DL models, the Ensemble-LSTM has established outstanding results and is considered a suitable

model for sleep stage classification concerning good f-measure, higher accuracy of 89.01%, and lower validation loss.

Table 6. Per-class performance achieved by E- LSTM Models on SSC Dataset

Sleep stage	Predicted on SSC dataset					Evaluation metrics (%)		
	W1	N1	N2	N3	REM	Precision	Recall	F-measure
W1	7726	206	42	32	21	96.25	96.58	96.19
N1	94	350	78	5	87	56.29	45.51	50.57
N2	90	104	3286	81	60	90.26	82.70	86.77
N3	76	24	140	1019	40	78.44	88.87	85.13
REM	60	146	163	32	1208	75.07	88.71	85.02
Overall	Accuracy=89.01 %					Kappa=0.838		

Table 6. shows the per-class performance achieved by the ensemble LSTM model for the sleep stage dataset (single channel). The diagonal values in the confusion matrix represent True Positive (TP) and imply that the number of sleep stages is correctly classified. The table shows the classification performance of each stage in terms of precision, recall, overall accuracy, kappa, and f-score. The model performs better for stages W, N2, N3, and REM, except for N1. It may be because the N1 stages have fewer epochs than the other stages. However, ensemble LSTM achieved better performance when compared with other state-of-the-art models (cascaded, Elman, attentional RNN) except for the N1 stage, as shown in Table 8. The reason is that other models classified sleep stages using fewer sleep stage epochs. The kappa (k) values showed a significant level of agreement between the E-LSTM model and the sleep expert.

Table 8. Comparative Accuracy analysis of the proposed E-LSTM with existing models

Models	Overall Metrics			Per-class F-Score				
	Sleep stage total (Epochs)	Accuracy (%)	kappa	W	S1	S2	S3	REM
Attentional RNN	-	79.1	0.762	75.5	27.3	86.6	85.60	74.8
Elman	2880	87.20	-	70.8	36.7	97.3	89.70	89.5
Cascaded	10280	86.74	0.79	95.29	61.09	85.48	84.80	83.74
Proposed E-LSTM	15170	89.01	0.838	96.19	50.57	86.77	85.13	85.02

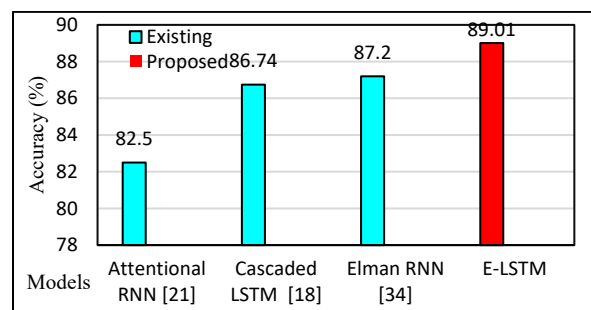


Figure. 8: Accuracy Analysis of the proposed E-LSTM with existing models

Table 8 shows a brief comparison of the ensemble models' results with those of existing models. In terms of accuracy, Figure 8 compares the proposed E-LSTM model to existing models.

Using the SleepEDF -18 dataset, the attentional RNN model, cascaded, and Elman RNN were used in the literature for SSC. The proposed ensemble LSTM model's performance is compared with that of the existing model, and the results are reported in Table 8. The results show that the attention mechanism has

accomplished a lower accuracy of 79.10%. Cascaded and Elman RNN models have obtained moderate accuracy of 86.74% and 87.20% [34], respectively.

As previously mentioned, it is evident that the Ensemble LSTM model outperforms the other model on the SSC. The experimental result reported that the ensemble LSTM had attained a higher classification accuracy with a good F-score. Hence, the performance of the E-LSTM model for SSC is observed to be a better model than other models reported in the literature.

## 6. Conclusion

This paper has effectively designed an ensemble of voting-based DL models with Regularization functions for sleep stage classification. At the initial stage, the input EEG data is pre-processed in stages such as channel extraction, feature extraction, and data normalization. Subsequently, three DL models, namely the RNN, LSTM, and GRU models, are employed for the classification of EEG sleep stages. A comprehensive set of simulations was done to validate the effective sleep stage classification outcome of the presented model, highlighting the superior outcome of the presented model. The obtained experimental results highlighted the improvement of the presented model on the test EEG sleep state dataset. While training the applied DL models, activity Regularization techniques are used to mitigate the overfitting problem. The proposed model overcame the overfitting problem that affected the model's performance. The DL model with activation Regularization techniques was used to close the gap between validation and training loss, which improved the model's performance. The max voting technique is used to determine the optimal SSC efficiency of the presented model. The experimental results showed that the ensemble RNN, ensemble GRU, and ensemble LSTM models had achieved an accuracy of 85.57%, 87.41%, and 89.01%, respectively, for sleep stage classification. In the future, bio-inspired optimization algorithms can be employed to determine the optimal weights in the voting technique. Additionally, the sleep stage is a sequential time series of various sleep stages (sub-bands), so one stage depends on the previous stage.

## Conflict of Interest

The authors declare no conflict of interest.

## References

- [1] A.D. Laposky, J. Bass, A. Kohsaka, F.W. Turek, Sleep and circadian rhythms: Key components in the regulation of energy metabolism, *FEBS Letters*, **582**(1), 142–151, 2008, doi:10.1016/j.febslet.2007.06.079.
- [2] J.C. Carter, J.E. Wrede, Overview of sleep and sleep disorders in infancy and childhood, *Pediatric Annals*, **46**(4), e133–e138, 2017, doi:10.3928/19382359-20170316-02.
- [3] S. Stranges, W. Tigbe, F.X. Gómez-Olivé, M. Thorogood, N.B. Kandala, "Sleep problems: An emerging global epidemic? Findings from the INDEPTH WHO-SAGE study among more than 40,000 older adults from 8 countries across Africa and Asia," *Sleep*, **35**(8), 1173–1181, 2012, doi:10.5665/sleep.2012.
- [4] F. Mendonça, S.S. Mostafa, F. Morgado-Dias, J.L. Navarro-Mesa, G. Juliá-Serdá, A.G. Ravelo-García, "A portable wireless device based on oximetry for sleep apnea detection," *Computing*, **100**(11), 1203–1219, 2018, doi:10.1007/s00607-018-0624-7.
- [5] Z. Roshan Zamir, N. Sukhorukova, H. Amiel, A. Ugon, C. Philippe, Optimization-based features extraction for K-complex detection, 2013.
- [6] L. Wei, Y. Lin, J. Wang, Y. Ma, "Time-frequency convolutional neural network for automatic sleep stage classification based on single-channel EEG," in *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, IEEE Computer Society: 88–95, 2018, doi:10.1109/ICTAI.2017.00025.
- [7] D. Wang, D. Ren, K. Li, Y. Feng, D. Ma, X. Yan, G. Wang, "Epileptic seizure detection in long-term EEG recordings by using wavelet-based directed transfer function," *IEEE Transactions on Biomedical Engineering*, **65**(11), 2591–2599, 2018, doi:10.1109/TBME.2018.2809798.
- [8] A. Ramachandran, A. Karuppiah, A survey on recent advances in machine learning based sleep apnea detection systems, *Healthcare (Switzerland)*, **9**(7), 2021, doi:10.3390/healthcare9070914.
- [9] T. Zhu, W. Luo, F. Yu, "Convolution-and attention-based neural network for automated sleep stage classification," *International Journal of Environmental Research and Public Health*, **17**(11), 1–13, 2020, doi:10.3390/ijerph17114152.
- [10] P. Jadhav, G. Rajguru, D. Datta, S. Mukhopadhyay, "Automatic sleep stage classification using time-frequency images of CWT and transfer learning using convolution neural network," *Biocybernetics and Biomedical Engineering*, **40**(1), 494–504, 2020, doi:10.1016/j.bbe.2020.01.010.
- [11] J. Zhang, R. Yao, W. Ge, J. Gao, "Orthogonal convolutional neural networks for automatic sleep stage classification based on single-channel EEG," *Computer Methods and Programs in Biomedicine*, **183**, 2020, doi:10.1016/j.cmpb.2019.105089.
- [12] I.N. Wang, C.H. Lee, H.J. Kim, H. Kim, D.J. Kim, "An Ensemble Deep Learning Approach for Sleep Stage Classification via Single-channel EEG and EOG," in *International Conference on ICT Convergence*, IEEE Computer Society: 394–398, 2020, doi:10.1109/ICTC49870.2020.9289335.
- [13] R. Boostani, F. Karimzadeh, M. Torabi-Nami, A Comparative Review on Sleep Stage Classification Methods in Patients and healthy Individuals A Comparative Review on Sleep Stage Classification Methods in Patients and healthy Individuals.
- [14] R. Sharma, R.B. Pachori, A. Upadhyay, "Automatic sleep stages classification based on iterative filtering of electroencephalogram signals," *Neural Computing and Applications*, **28**(10), 2959–2978, 2017, doi:10.1007/s00521-017-2919-6.
- [15] M. Sharma, D. Goyal, P. v. Achuth, U.R. Acharya, "An accurate sleep stages classification system using a new class of optimally time-frequency localized three-band wavelet filter bank," *Computers in Biology and Medicine*, **98**, 58–75, 2018, doi:10.1016/j.compbiomed.2018.04.025.
- [16] S. Chambon, M. Galtier, P. Arnal, G. Wainrib, A. Gramfort, "A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series," 2017.
- [17] P. Piñero, P. García, L. Arco, A. Álvarez, M.M. García, R. Bonal, "Sleep stage classification using fuzzy sets and machine learning techniques," *Neurocomputing*, **58–60**, 1137–1143, 2004, doi:10.1016/j.neucom.2004.01.178.
- [18] N. Michielli, U.R. Acharya, F. Molinari, "Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals," *Computers in Biology and Medicine*, **106**, 71–81, 2019, doi:10.1016/j.compbiomed.2019.01.013.
- [19] A. Supratak, H. Dong, C. Wu, Y. Guo, "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **25**(11), 1998–2008, 2017, doi:10.1109/TNSRE.2017.2721116.
- [20] H. Dong, A. Supratak, W. Pan, C. Wu, P.M. Matthews, Y. Guo, "Mixed Neural Network Approach for Temporal Sleep Stage Classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **26**(2), 324–333, 2018, doi:10.1109/TNSRE.2017.2733220.
- [21] H. Phan, F. Andreotti, N. Cooray, O.Y. Chén, M. de Vos, Automatic Sleep Stage Classification Using Single-Channel EEG: Learning Sequential Features with Attention-Based Recurrent Neural Networks, 2018, doi:10.0/Linux-x86\_64.
- [22] Y. Yang, X. Zheng, F. Yuan, "A study on automatic sleep stage classification based on CNN-LSTM," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, 2018, doi:10.1145/3265689.3265693.

- [23] S. Mousavi, F. Afghah, U. Rajendra Acharya, "Sleeppegnet: Automated sleep stage scoring with sequence to sequence deep learning approach," *PLoS ONE*, **14**(5), 2019, doi:10.1371/JOURNAL.PONE.0216456.
- [24] M.J. Hasan, D. Shon, K. Im, H.K. Choi, D.S. Yoo, J.M. Kim, "Sleep state classification using power spectral density and residual neural network with multichannel EEG signals," *Applied Sciences (Switzerland)*, **10**(21), 1–13, 2020, doi:10.3390/app10217639.
- [25] M. Abdollahpour, T.Y. Rezaei, A. Farzammia, I. Saad, "Transfer Learning Convolutional Neural Network for Sleep Stage Classification Using Two-Stage Data Fusion Framework," *IEEE Access*, **8**, 180618–180632, 2020, doi:10.1109/ACCESS.2020.3027289.
- [26] S. Hochreiter, *Recurrent Neural Net Learning and Vanishing Gradient*, 1998.
- [27] G. van Houdt, C. Mosquera, G. Nápoles, "A review on the long short-term memory model," *Artificial Intelligence Review*, **53**(8), 5929–5955, 2020, doi:10.1007/s10462-020-09838-1.
- [28] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," 2014.
- [29] S. Salman, X. Liu, "Overfitting Mechanism and Avoidance in Deep Neural Networks," 2019.
- [30] X. Ying, "An Overview of Overfitting and its Solutions," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, 2019, doi:10.1088/1742-6596/1168/2/022022.
- [31] W. Qingjie, W. Wenbin, "Research on image retrieval using deep convolutional neural network combining L1 regularization and PRelu activation function," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics Publishing, 2017, doi:10.1088/1755-1315/69/1/012156.
- [32] S. Merity, B. McCann, R. Socher, "Revisiting Activation Regularization for Language RNNs," 2017.
- [33] A. Dogan, D. Birant, "A Weighted Majority Voting Ensemble Approach for Classification."
- [34] Y.L. Hsu, Y.T. Yang, J.S. Wang, C.Y. Hsu, "Automatic sleep stage recurrent neural classifier using energy features of EEG signals," *Neurocomputing*, **104**, 105–114, 2013, doi:10.1016/j.neucom.2012.11.003.



## Integrated GIS-SUE Map Cost Estimation System Prototype for Designing a Decision Support System

Ali Nashwan\*, Khalil Al-Joburi

Civil Engineering, University of Bahrain, Sanad, 00745, Bahrain

### ARTICLE INFO

Article history:

Received: 29 August, 2022

Accepted: 07 January, 2023

Online: 07 February, 2023

Keywords:

GIS

SUE

Smart cities

UML

Utilities

### ABSTRACT

*Subsurface Utility Engineering (SUE) is an international model for mapping and classifying underground surfaces according to their accuracy (acquisition method). Utilizing Geographic Information System (GIS) to map and present the SUE levels paved the way for producing a new Decision Support System (DSS) for the utility mapping process. The proposed system represents an efficient tool for managing, operating and maintaining utilities. This Article aims to design a prototype in Unified Modeling Language (UML) of a new DSS system to operate SUE maps using digital spatial maps (GIS-compatible). Although SUE and GIS are not new technologies, integrating them is. The result is a prototype that makes utility management and maintenance cost estimation more efficient. This prototype facilitates and automates the cost estimation of exposing, maintaining, or locating subsurface objects, such as utilities. In addition, it may apply to Municipal Solid Waste (MSW) and void mapping.*

### 1. Introduction

This paper is an extension of work originally presented at the 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT) [1].

Planning is a basic management function involving formulating one or more detailed plans to achieve an optimum balance of needs or demands with available resources (Source: <https://www.dictionary.com>). The cost is the most vital function to consider during the planning process. On the other hand, the cost usually changes from one location to another. Thus, spatial analysis is considered one of the best cost estimation and selection tools. This geospatial technique is called Location Based Service (LBS). LBS accuracy depends on the mapping accuracy of spatial features. Therefore, accurate LBS produces accurate cost estimation.

Spatial analysis is one of the main functions of GIS. However, its power and efficiency are inherently in the ability to combine spatial with non-spatial data called "Attributes". Figure 1 shows an example of a GIS map with an attribute table (Geodatabase).

The GIS in utility discussed in Article [1] started by illustrating the Facility Mapping systems (FM). These systems enhance the construction, maintenance, and operation of utilities. FM system mainly represented the utility features as vector (point, line, or

polygon) rather than raster data (pixels). This is because vector data has better accuracy than raster and is easy to calculate.

Increasing mapping accuracy of the existing subsurface utility minimizes the potential risk of damaging them. This objective is significant for the majority of subsurface mapping researchers. Therefore, SUE was designed to control and document underground utilities' mapping accuracy. In addition, SUE has several other benefits illustrated in [1].

SUE classifies the mapping features based on source accuracy (Data capturing). The classification is usually divided into four categories (Levels of Quality). These categories are class D: the location was determined based on historical data (digitizing old drawings or asking experts). Class C indicates that a topographical survey was used to locate the existing surface manholes and utility markers. Class B means that a geophysical surveying method was conducted to locate the utility, such as Ground Penetrating Radar (GPR), Electricity Pipe and Cable Locator (EPL), and other techniques. Finally, Class A refers to a non-destructive drilling technique (using water or air jetting devices) or hand-work trial holes implemented to identify the locations. The reliability of these classes was set to 25, 50, 75, and 100% for the classes D, C, B, and A, respectively [2]. Figure 2 shows a typical SUE map with the level of accuracy (confidence) and depth for each feature written over it.

These levels from A to D are inversely proportional to cost and directly proportional to risk, as shown in Figure 3.

\* Corresponding Author: Ali Nashwan, Bahrain - Sanad, 00745, +97339886460  
[alinashwan@gmail.com](mailto:alinashwan@gmail.com)

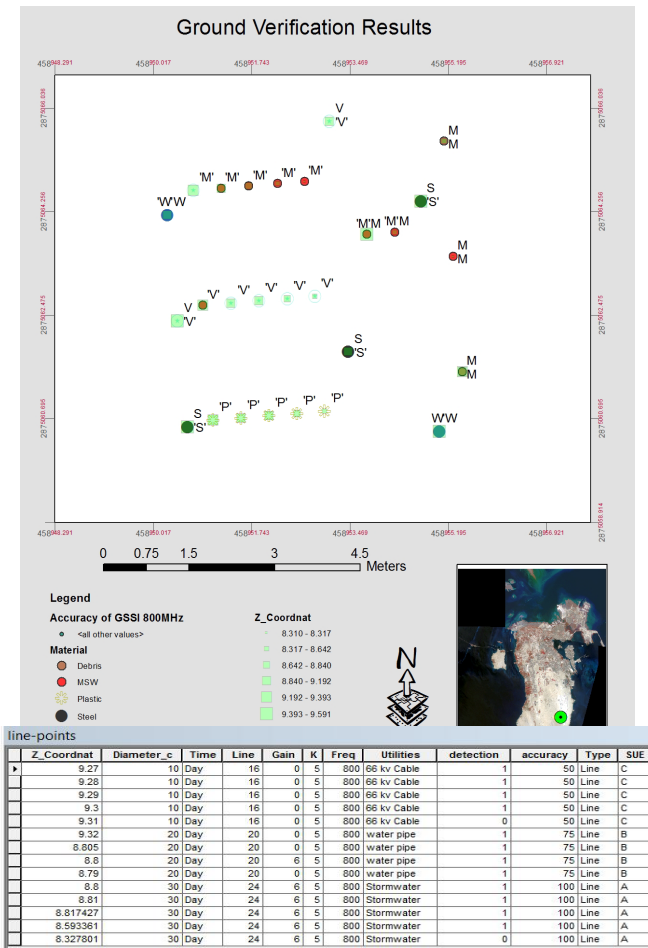


Figure 1: Sample of GIS map with attribute data

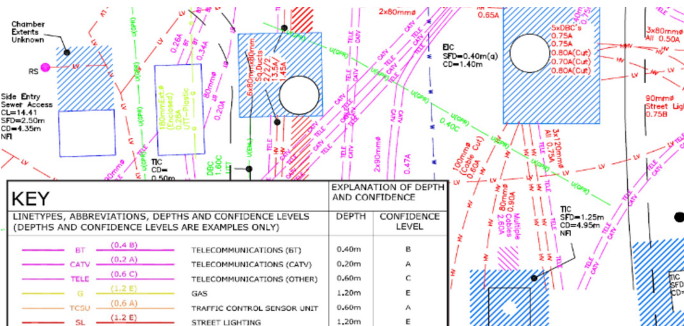


Figure 2: Typical SUE Map

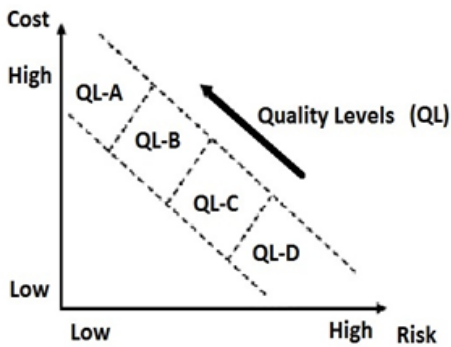


Figure 3: SUE Quality levels

Article [1] discusses integrating GIS with SUE. Also, the Article provided a prototype for this integration (How to convert the SUE map into the FM system and how to add the SUE classes to the FM system).

The future work for the mentioned Article was to design a DSS that helps the FM system user estimate the cost and time for mapping subsurface utilities. Therefore, this research extends this work to produce the mentioned DSS. In addition, this Article and [1] represent an enhancement of the results obtained by the original research in subsurface mapping [3].

Several works of literature define and illustrate the SUE and GIS in more detail, such as the articles [4,5], and [6], which represent good references that discuss SUE's benefits and history. These articles showed that the SUE has begun enhancing utility mapping certainty in roads and highway projects. They also show that the benefit of investment in SUE is enhanced and increases with time.

On the other hand, the articles [7,8], and [9] discuss the FM and the importance of GIS and FM in managing utilities and other subsurface features. In addition, these articles show that FM represents an office automation system that starts from converting the utility maps from CAD to GIS layers and includes utility planning, designing, operating, maintenance, and construction.

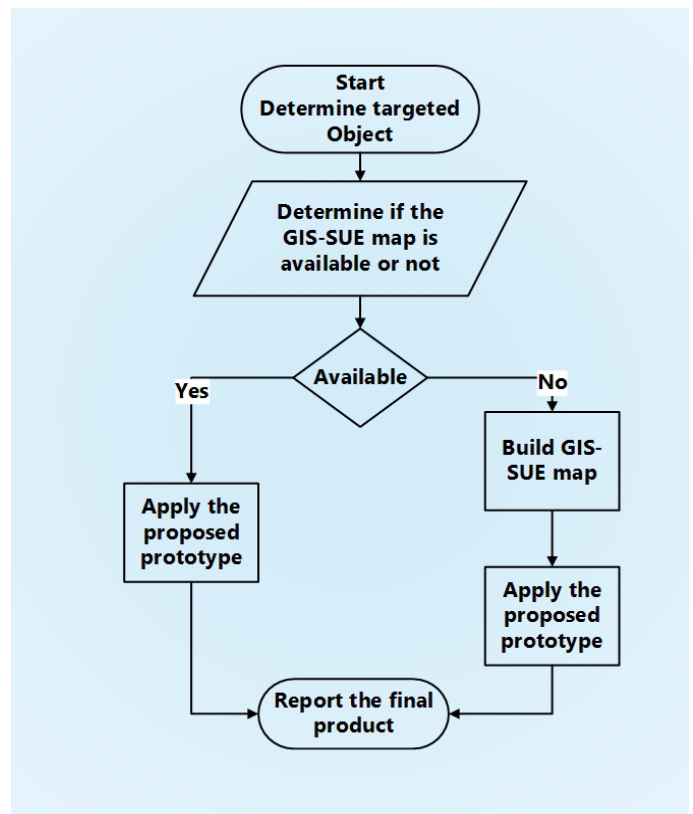


Figure 4: Methodology

## 2. Methodology

This research aims to design a prototype for a DSS that enhances the cost and time estimation for subsurface feature mappings, such as utility construction and utility data acquisition. The proposed system is designed based on the prototype for integrating GIS data and SUE maps.

The methodology of this research flowchart is shown in Figure 4, started by defining the targeted object to be investigated. The targeted objects in this research are utilities, voids, and MSW because the original researches [1] and [3] were designed to study these objects. Even though the outcome prototype can be applied to investigate any subsurface features, the system was built based on the outcome of the integration and conversion between SUE and GIS layers. Therefore, the availability of a GIS-SUE map might be defined to delineate the scenario that the software (system) will follow (either to build a new GIS-SUE map or to work on an existing one). Thus, this research focuses on putting the designed prototype into a system that determines how much it costs to locate the intended feature and to gauge its cost and time.

This Article presents the significance of utilizing SUE in the proposed DSS. Then, briefly discusses the designed prototype. Lastly, it discusses the results and output of this research.

### 3. SUE in a Decision Support System

The benefit of SUE levels could save the cost of data acquisition, depending on data acquisition objectives. For example, there is no need to pay a significant amount to gather data using topographical or geophysical surveying; when the map is for planning purposes, historical data is enough. Also, it assists the project manager in determining the requirements for the project. For example, if the map is for construction purposes, the area of potential risks requires level A (precise location) data, such as junctions, to avoid hazards and injuries.

The previous examples are used to establish a DSS to help the designer and the project manager determine the cost and risk of utilizing the feature's locations in further procedures, such as excavation and construction. However, the proposed DSS depends on the pre-requisite information provided by the user. This information helps the system define the likelihood, track and calculate the costs. The requirements are:

- 1- The objective of the requested mapping is to delineate the required level of SUE.
- 2- SUE level for the existing feature, if any.
- 3- The unit price for locating the service includes the exposing method, the depth, length, size, action required (protect, relocate, or do nothing), and any other information that affects the total utility cost.
- 4- The time required to implement the action in the previous step.

The proposed system will delineate the cost of the required operation near the utility feature with a range of certainty that depends on the level of SUE of the original data. In addition, the system estimates the cost of changing the level of SUE data to increase the reliability of the data, which must be done later in the construction phase.

The proposed prototype shown in Figure 5 is designed for utility. However, it could be implemented and updated for other targets' inspections, such as MSW and voids, where the requirements differ. For example, the inspection cost might be included if this DSS is used in MSW mapping. However, contrariwise, the utility's locations are usually known with

confidence (the inspection and the SUE ensure these locations) voids and MSW locations are usually unknown. Thus, the inspection is done to locate or minimize the suspected areas. Then additional inspections or excavations are done to locate it. In addition, the location of utilities is required to maintain and protect them, while the MSW and voids location must be determined to remove or treat these features. Therefore, it is recommended to start with the inspection cost, locate the features, and remove them to avoid collapse and cracks that could deteriorate the surface structure, like the collapse due to a void or the cracks due to MSW under the building foundation.

The following example is to clarify how the cost and time estimation of locating the objected features (utility, voids or MSW)

will be calculated in the system:

The pre-requisite items for the system are the GIS-SUE map for the area of interest for objected features. In addition, metadata (extra data) contains the amount of cost and time for locating each type of SUE level (C, B, or A), which is called a Bill of Quantity (BOQ). (Such as the cost of using the topographical survey to locate the service in level C, using the geophysical survey to locate it in level B, or exposing the location (trial hole or non-destructive digging) to obtain level A location.

The objective in the example is to find the cost and time for locating an object (water pipe) during the construction phase (which means locating the pipe in level A). If the existing level is C, then the location needs to be inspected using the geophysical method (GPR in this case) (to get level B type) and to do some trial holes for verification (for level A). Using the BOQ, the cost and time of both types could be calculated inside the system and added as an attribute on the feature or in a separate report.

### 4. Implementation

The system started by gathering the required information from the user, which was mentioned previously. This information started by determining the required SUE level for the targeted feature (as mentioned in the introduction. In addition, the system might determine the expected area to be investigated using the original GIS or CAD map. Then using the price, time, and quantity database, the cost will be determined. Lastly, the output of the DSS will be a GIS layer and a report in an MS Excel sheet.

The prototype shown in Figure 5 is working as follows:

---

**Result:** Determining the cost and time required to implement a specific process in the project.

#### **Initialization;**

Gather the project data from the user (existing SUE-levels, Unit price for time and cost for each item, and data mapping method).

Refine the collected information for the decision analysis process.

**Implementing:** the following **algorithm** is used to determine if the SUE level is suitable for the project's objective and to determine the required cost and time for it (how the system will work). The example for this algorithm is the calculation of the cost and time for the construction phase:

**While** data is not sufficient to calculate the cost & time, **do**

Apply the sub-prototype in Figure 6;  
 Save the cost and time in a database;  
 The following verifying process is used to check objective  
 With SUE level (data accuracy) using the sub-prototype:

**If Data accuracy is not at level A, then**  
 Calculate the time and cost for the required action  
 (exposing, inspection, relocating);

**else**  
 Do nothing,

**end**  
 The accuracy is enough for the objective, and it up  
 to 100% for level A,  
 end sub-prototype

**End**

Convert the results into a shapefile.

The outputs are a GIS map where the user can start the activity, he/she wants (such as construction) and the area that still has risk (an area that is lower than level A);

## 5. Results and Discussions

The proposed solution in this research was a prototype designed in UML high-level format. The complete programming project required a budget, further details, and available software. This prototype can be programmed and implemented partially using MATLAB code or any other programming language.

Although the main utilization of the proposed DSS might be in a subsurface utility map, it could be used to map any subsurface feature, such as MSW and voids. This opens the way to expand the DSS to be adopted in infrastructure applications for subsurface mapping purposes.

The SUE was considered a valuable tool in the damage-preventing process. Also, it helps in reducing utility relocating, project delays, and compensations. Besides, facilitates utility maintenance and several other benefits. Although the main item in the mentioned literature was the utilities, this Article investigates other inspections, locating and removing other potential risks such as MSW and voids after considering some changes in the DSS. Therefore, the primary investigation was designed in the original project mentioned in [3] to detect these features instead of utilities.

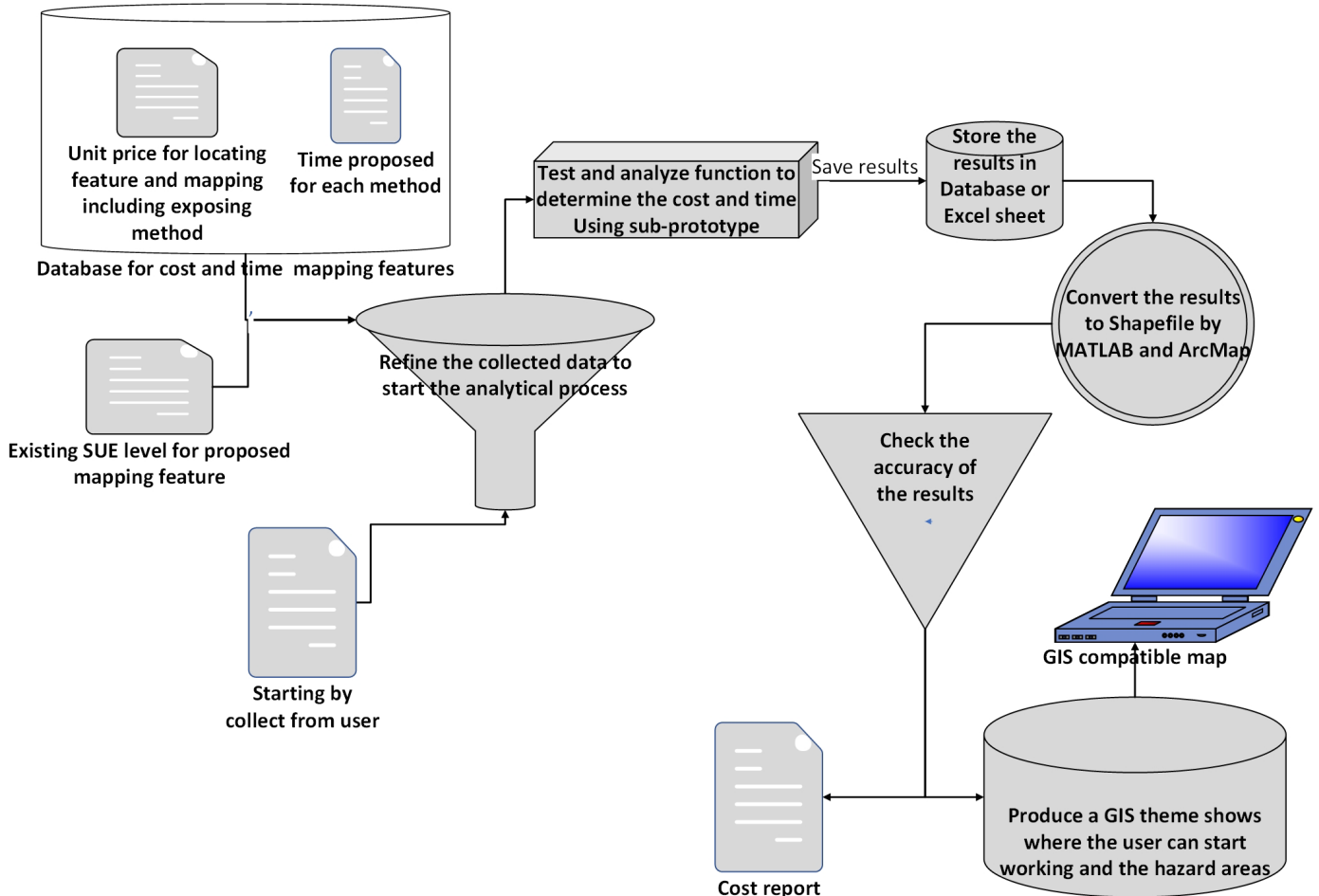


Figure 5: Proposed General DSS prototype for cost and estimation

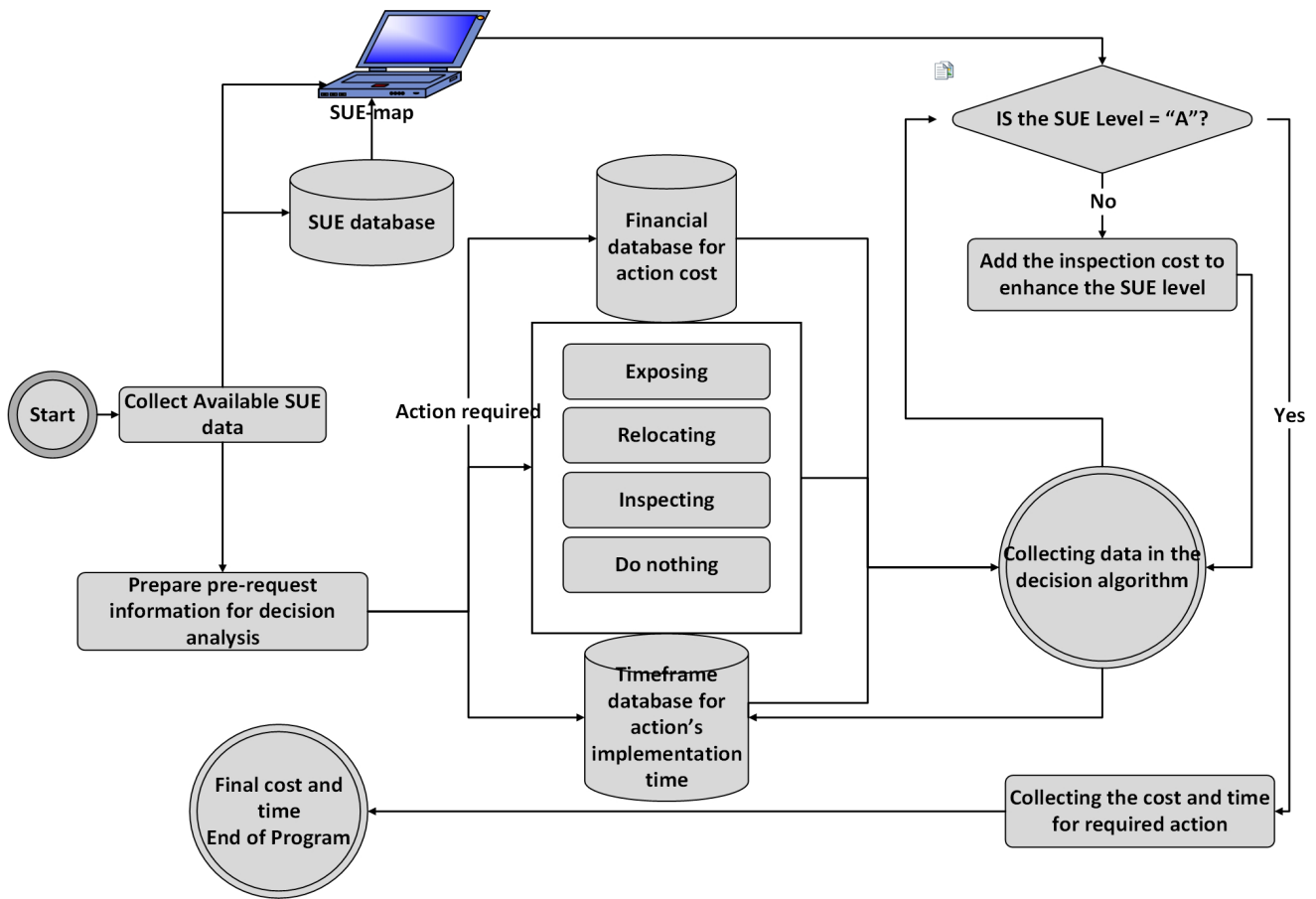


Figure 6: Sub-Prototype for DSS cost analysis in construction phase

The significance of accurately locating the other subsurface feature is to prevent the potential structural damage of these features, such as cracks and collapse due to voids. These locations are essential for structural design and affect the final cost of the construction project, such as exposing and treating these features or negating their effect with other structural solutions like reinforcement concrete mixes. On the other hand, the maintenance engineer might utilize the SUE map to delineate the expected time and cost to repair the utility faults. Therefore, he/she can decide which is feasible to expose the old utility or cancel it and lay a new one. In addition, the planner and structural engineers might change the proposed design to avoid underground MSW or voids or to add the cost of exposing and removing it before construction. Thus, the proposed prototype illustrated previously for a decision support system using an integrated GIS-SUE map was designed to solve this issue.

The results of the proposed DSS prototype are a GIS map beside a report

### 6. Conclusions

This Article presented a prototype to design and implement a DSS for determining the cost and time required to do a specific action on a utility network segment or locate subsurface hazardous material. This prototype was designed to analyse input data, such as the unit cost for each type of action with the targeted feature. The prototype compares the objectives to the available confidence level (SUE). The output is a GIS-compatible map and a report outlining the cost and time needed to upgrade the SUE certainty

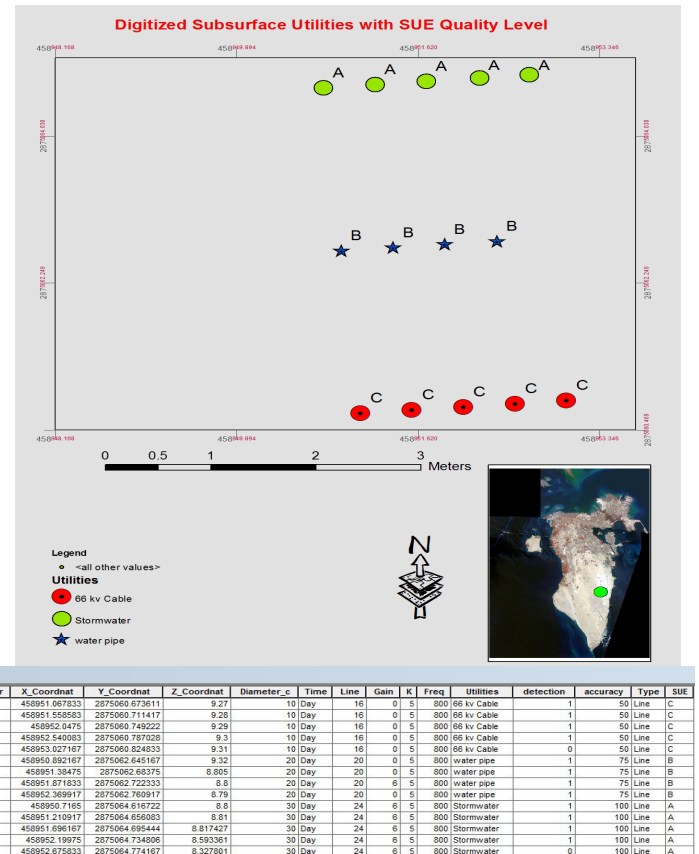


Figure 7: GIS-SUE integrated map with attribute

level. Such a system represents a new step in smart cities and office automation. In the past, the same procedure might have been done manually and by geophysical engineering experts.

The next step is to implement and test the prototype. The resulting maps provide several spatial analyses that help project managers, planners, and quantity surveyors assess the areas requiring an extra budget for upgrading data accuracy during construction and maintenance.

Maintenance costs for existing subsurface utilities can be estimated upon customizations and linked with utility maps.

## References

- [1] A. Nashwan, K. Al-Joburi, "A Prototype to Produce an Integrated GIS-SUE Map," in 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), IEEE, Bahrain: 592–597, 2021, doi:10.1109/3ICT53449.2021.9581853.
- [2] S. Rana, N. Swarup, SUE Training Module Level One, 56, 2018.
- [3] A. Nashwan, 3-D Subsurface Feature Mapping: Integrated Geospatial Approach, University of Bahrain, 2021.
- [4] Y.J. Jung, "Evaluation of subsurface utility engineering for highway projects : Benefit – cost analysis," *Tunnelling and Underground Space Technology*, **27**(1), 111–122, 2012, doi:10.1016/j.tust.2011.08.002.
- [5] B. Uslu, Y.J. Jung, S.K. Sinha, "Underground Utility Locating Technologies for Condition Assessment and Renewal Engineering of Water Pipeline Infrastructure Systems," *Journal of Pipeline Systems Engineering and Practice*, **7**(4), 221–242, 2016, doi:10.1061/(ASCE)PS.1949-1204.0000221.
- [6] S. Rana, Advance GPR Training Course: Data processing, 2020.
- [7] A. Fenais, S.T. Ariaratnam, S.K. Ayer, Nikolas Smilovsky, "Integrating Geographic Information Systems and Augmented Reality for Mapping Underground Utilities," *Infrastructures*, **4**(60), 1–17, 2019, doi:https://doi.org/10.3390/infrastructures4040060.
- [8] M. Wang, Y. Deng, J. Won, J.C.P. Cheng, "Automation in Construction An integrated underground utility management and decision support based on BIM and GIS," *Automation in Construction*, **107**, 1–22, 2019, doi:10.1016/j.autcon.2019.102931.
- [9] J.R. Meyers, "GIS in Utilities," *Geographical Information Systems: Management Issues and Applications*, **2**, 801–818, 1999.

## Conception and Simulation of an Electronic Nose Prototype for Olfactory Acquisition

Mostapha Harmouzi \*, Aziz Amari, Lhoussaine Masmoudi

Conception and Systems Laboratory, Faculty of Sciences, Mohammed V University in Rabat, Morocco

---

### ARTICLE INFO

Article history:

Received: 01 September, 2022

Accepted: 08 January, 2023

Online: 07 February, 2023

---

Keywords:

e-nose

Sensing chamber

Sensor array

Gas compartment

---

### ABSTRACT

The "Electronic Nose" approach, which is exclusive to gas measurement systems, uses gas sensors as odor detectors. Design faults exist in the existing electronic nose (e-nose) chamber, such as its large volume, difficult construction, etc. In order to obtain measurements in a satisfactory state, we want to create a gas chamber that can provide favorable conditions for the sensor array, taking into account the ideal gas flow morphology and detector placement. To describe and identify the design capable of offering the best performance for a genuine idea, the e-nose chamber was created using ParaVIEW simulation and FreeCAD conception. According to the results, the spherical sensing container with connections from both pipes in a tangential arc style gives the highest performance in terms of turbulence reduction, in that case, we are printing this chamber and put it in a gas flow prototype to see the performance of the quality measurement of the sensors inside it, and the result shows that these sensors have good acquisition responses by testing the homogeneity distribution inside the chamber.

---

## 1. Introduction

The human olfactory system, which consists of the region from the olfactory epithelium to the olfactory brain, is a sensory system for the detection of odor molecules. Olfaction is the term for the sensory function of smell that allows humans to detect and identify a wide range of odorants. Odorants enter the mucus-covered olfactory epithelium in the nasal cavity, bind to the olfactory receptors present in the cilia of the olfactory sensory neurons, and then send odorant information to the brain [1].

Odorants are volatile, hydrophobic compounds that have molecular weights of less than 300 Daltons. The largest known odorant to date is labdane that has a molecular weight of 296 [2].

The sense of smell is a very sensitive organ that can detect even the smallest amounts of substances. It is estimated that only 2 % of the volatile compounds available in a single sniff will reach the olfactory receptors, and as few as 40 molecules of some mercaptans are sufficient to perceive an odor [3].

Instead than being caused by a single chemical, most odors are created by combining hundreds of different odorants. Perceptual fusion results from the tendency of individual components to harmonize or blend together in mixes. The maximum number of

components that humans can detect in a combination of odors is three to four [3].

A system that addresses the demand for increasing production or quality control against adulterations is crucial in the fields of food and agricultural product quality control [4]. The most crucial component of the overall e-nose system is the sensor array, which responds differently to different gas molecules [5]. The detecting system of an e-nose is composed of a sensing element, chamber, and sensor array. In order to maximize interaction between the sensor array and volatile substances, the sensing chamber can be a dynamic closed-loop area. The sensing chamber may also be a static area that prevents gas leaks during the detecting stage, which might have a negative impact on the interpretation of the data [6].

One of the biggest issues with an e-detection nose's stage efficiency is poor sensor response to volatile chemicals. One of the most frequent issues related to subpar sensor response signal performance is also thought to be the longer time it takes for sensors to attain steady state.

The size and shape of the detecting chamber, together with the placement of the intake and outflow gases, are all factors that might significantly affect the performance of the sensor's response. The gas flow and concentration in the detecting chamber are impacted by these variables.

---

\* Corresponding Author: Mostapha Harmouzi, Email: [mostapha7.harmouzi@gmail.com](mailto:mostapha7.harmouzi@gmail.com)

## 2. Technology for an e-nose system

An e-nose system is a device with a sensor array that is intended to distinguish between and identify complex scents. Three primary components make up e-nose instruments: a sample handling system, a detection system or sensor array, and data processing models, including categorization and prediction models. The sensor array consists of general-purpose sensors, often metal oxides, that have undergone various chemical treatments. When the volatile molecules are exposed to the sensor array, a particular smell print (or fingerprint) is instantaneously produced from them. To build the database and train a new pattern recognition system that is used to categorize and identify novel scents, patterns or fingerprints from known odors are employed [7].

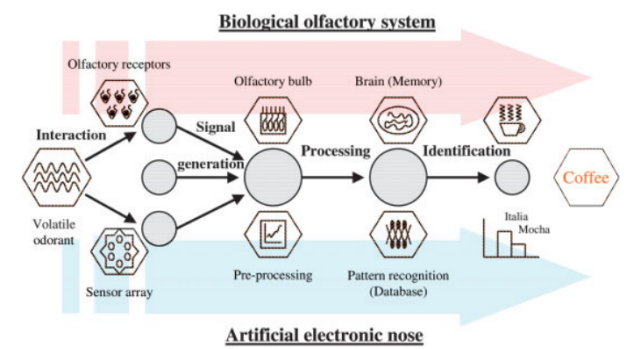


Figure 1: Basic diagram showing the analogy between biological and artificial noses [8].

Electronic nose is a quick and powerful approach that does not require any specific sample preparation and can determine a product's whole volatile profile. Devices called e-noses are composed of several sensors and may mimic the sense of smell.

As shown by the similarity between biological and artificial noses in Fig. 1, e-nose is designed as a match-model for the natural nose and includes the many phases between a volatile odorant and its detection, namely: interaction, signal generation, processing, and identification. A sensor array, electrical interface circuitry, and a pattern-recognition device that serves as a signal processing system are all included in the system. However, a more straightforward model based on a collection of sensors, signal amplification through pre-processing [9], and a pattern recognition system aids in better understanding and illustrating how the nose works [8]. Pattern recognition methods include all statistical and neural techniques for the classification and recognition of odors.

Machine learning is used in the electronic nose. The core concept is that the measurement system learns from training measurements to achieve the desired goal rather than the scientist or application engineer adapting the measurement system to a specific measurement activity [6].

### 2.1. Measurement conditions stability

The electronic nose's seeming simplicity has made processing samples and delivering sample gas easier. Furthermore, despite the fact that many electronic nose scientists have academic

backgrounds in mathematics or electronics, they lack knowledge of chemical-analytical concerns. The same is true for the measurement systems. Data from a basic electronic nose measuring system, such as one that merely consists of a few gas sensors in a straightforward chamber, will be noisy and have a high degree of measurement uncertainty [6].

Therefore, the only means of achieving stability are controlled sampling methods and the transfer of volatiles to the measurement apparatus. The electronic nose must also be as technologically advanced as is feasible to block sources of noise and outside impacts.

### 2.2. Detection measurement

The three main parts of an electronic nose are a detecting system sample, a processing system, and a calculation system. The sample handling technique enables the generation of a sample's headspace (volatile chemical smells [6]). The detecting component of the electronic nose is then injected with this headspace. The sample handling method is essential for ensuring constant operational conditions [10].

Our aim study is based on the detection system sample, because firstly before entering to the processing and computing phases we should make a good measurement from the sensor array in e-nose system by using the procedure below [4] (Figure 2):

- To suppress any strange gases in the cell, first of all to do is injecting the gas vector (oxygen, nitrogen, or air) into the chamber, which is controlled by opening the upper electrical valve (the strange gases are considered as noise in the measurement).
- When to confirm that the chamber is cleaned, then flowing the gas vector into the sample to through the chamber sensor array by opening the downer electro-valve and closing the first one.
- Recording the data from any acquisition electronic board and plotted to see the measurement of the sensor array.

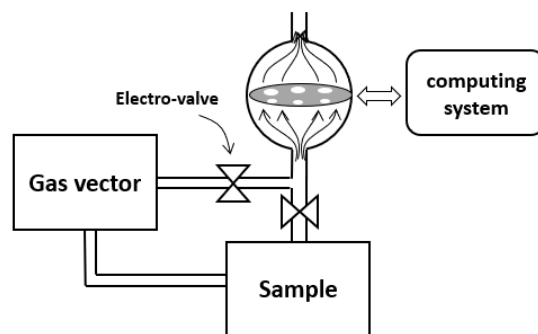


Figure 2: Procedure sensing of an e-nose system [4].

## 3. Prototype system design

In order to make sure that gas sensors offer reliable readings, we look at the process design of the sensing chamber and the notion of hydrodynamic and fluid flow in this section [11]. We will assess the form, taking into account the sizes depend on the number of sensors employed, and temperature and humidity impact.



### 3.1. Specification

Two fluid systems types are, static and dynamic systems. According to their names, static systems are those in which the fluid is at rest and dynamic systems are those in which the fluid is in motion. [12].

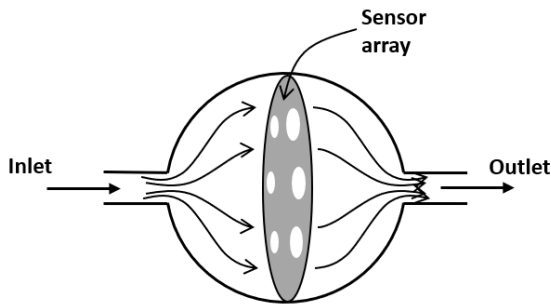


Figure 3: Sensor array inside the gas chamber of e-nose [4].

A dynamic system maintains gas concentrations that are utilized for processing. As seen in Figure 3, the chamber's design should allow for a more effective dispersion of the gas introduced within. To do this, a variety of sensors with the ability to simulate fragrance must be used [13].

### 3.2. Chamber design

With a single inlet and outflow, the chamber was meant to resemble a spherical chamber (Figure 3). This shape may be compared to another in which the detecting system performed poorly because the stationary zone was big, for instance because the rectangular shape's borders trapped the gas (Figure 4). As a result, the chamber's sensitivity to compost emissions was poor, and its signal-to-noise ratio was less stable [5].

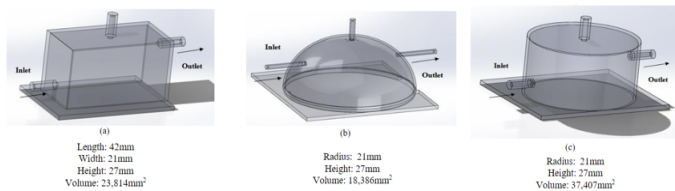


Figure 4: Shapes and volumes of the (a) rectangular, (b) hemisphere and (c) cylindrical sensing chambers [5].

In [5], the author concluded from a comparison of various chambers that the hemisphere chamber, shown in Figure 4, which they used to assess the sensor's time response and stagnant region, had the best performance. For each sensor chamber and the alcohol content, the average of three readings was obtained in 100 seconds [5]. When there was just clean air in the sensing chambers at first, the alcohol content remained consistent for all of them. The three sensing chambers show a rapid rise in alcohol concentration level with passing time, however the rectangle chamber responds more slowly than the hemisphere and cylindrical chambers when they inject the alcohol gas detected by the gas sensor (MQ-3) inside the sensing chambers [5].

However, the author study [5] encouraged us to specify the e-nose geometrically as a spherical shape, which we base on the temporal

response and additional measuring benefits, as illustrated in Figure 5.

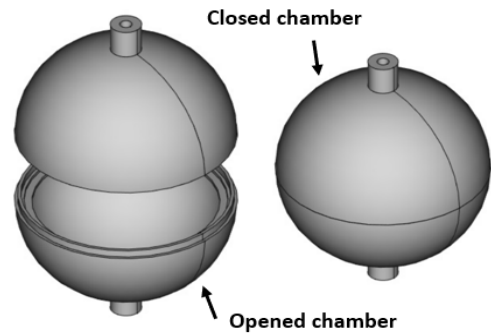


Figure 5: Spherical chamber opened and closed[4].

### 3.3. Sensor array arrangement

Odors of interest in practical applications are typically complicated mixes rather than pure gases. A fruit's aroma, for instance, is a sophisticated fusion of dozens of distinct odors. Finding sensors for each individual component of the gas mixture is very hard due to its complexity [14].

In the literature we found that, six to seven gas sensors are employed in the nasal sensor array [15]-[16]-[17], some researchers supporting them to eleven or more [18]-[19]. In our case, six sensors used for the measurement, which are useful to describing the odor of the sample. To be more practical, addition of temperature and humidity sensors is advised. (DHT11), which are important in the measurement and signal processing [4].

Our sensor array should be built on a PCB circle shaped that fits our spherical chamber. The hexagonal dispersion geometry makes a symmetrical form of these sensors as seen in Figure 6.

Temperature has an impact on other aspects of thermal comfort and indoor air quality [20]. The temperature in the occupied space ought to stay constant. On the other side, high humidity can promote the development and proliferation of molds and bacteria that can spread via the air. Controlling quality is a difficulty in this situation. On the other hand, relative humidity has a big role in thermal comfort. The optimum range for the relative humidity of indoor air, as per international norms and regulations, is between 30 and 60% [21].

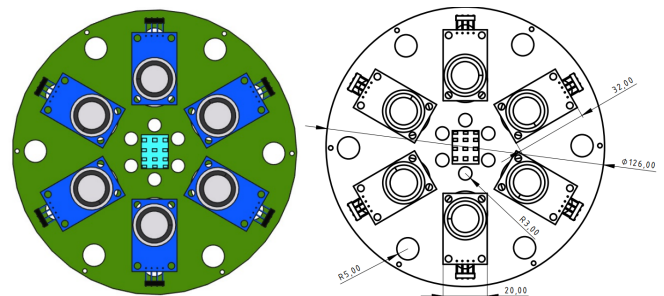


Figure 6: Sensor array network distribution in hexagonal form, as well as temperature and humidity sensors.

### 3.4. Flow rate

The time needed to reach a uniform condition and the chamber's capacity were two of the many variables used to evaluate the sensor's response signal performance. As shown in Figure 7, the design of a spherical chamber performed well since the stationary zone was minimal and the chamber was less intricately constructed.

The continuity equation, a fundamental principle of fluid mechanics [11], asserts that the volume flow is constant and that the quantity of any incompressible amount of fluid entering and departing are equal, The measurement item and the chamber volume determine the flow rate. Figure 7 depicts the geometry of the cell.

The calculation of velocity and flow rate, provide us to know the volume of chamber designed by applied some formulas.

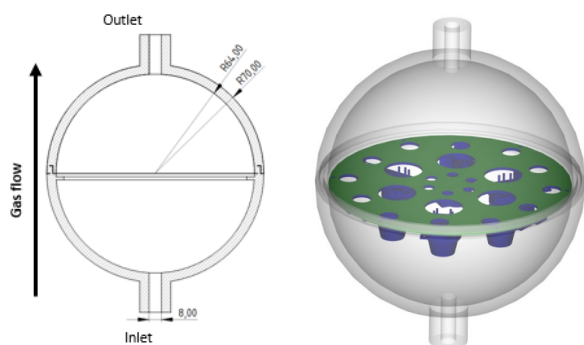


Figure 7: Dimension of the e-nose shape (mm unit).

$$V = \frac{4}{3} * \pi * r^3 \quad (1)$$

$$S = r * \pi * 2 \quad (2)$$

$$D = S * v \quad (3)$$

where:

V = Volume of the chamber

S = Section of the tube

D = Debit (flow rate)

v = velocity

r = Radius

For the velocity, our object is to make a symmetrical concentration of the gas inside the sensing chamber in a good time, fixed about 180 seconds [16], and the inlet and outlet pipe section is 50.27 mm<sup>2</sup> by (2), the volume of the sensing chamber using (1) is 1098 cm<sup>3</sup>, and velocity equal to 3.98 cm/s. With these parameters we can calculate the optimum flow rate 2 cm<sup>3</sup>/s using (3).

### 3.5. Architectural prototype system for sensing chamber test

The prototype's main objective is to control the quality of agricultural products by using smell (gas sensor matrix) through the first step to ensure the reliability of the information as much as

possible. Performed by a procedure, elaborated in the next section, well determined at the fluidic level, to acquire the necessary data for the processing system.

In general, the experiment takes place in a well-defined time (according to the response of the gas sensor), at first the odorant is putted in the sample door, and we wait for any moment that the smell is concentrated before the flow. Then any smell is released inside the gas cell with the air by the action of opening the solenoid valve at the top V1 plus the start of the pump until the stabilization of the sensors as shown in Figure 8. The next step is to open the solenoid valve at the bottom V2 and close the solenoid valve at the top V1, so that the smell of the sample flows to the gas cell, which causes a measurement exchange at the sensor level.



Figure 8: Prototype system conception.

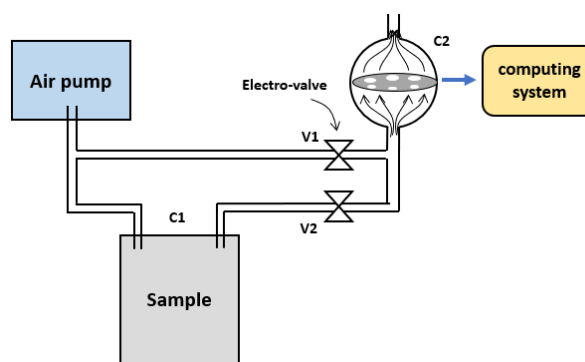


Figure 9: Synoptic diagram.

## 4. Results and discussions

### 4.1. Chamber simulation

Working on FreeCAD software using the workbench CfdOF with OpenFOAM tool to simulate the sensing chamber by setting some parameters as, ambient temperature, boundary conditions, iterations, the temperature, density and flow rate. And importing the mesh result to ParaVIEW software. According to Figure 9, we visualize the streamline movement.

In fact, energy is wasted at the level of gas movement (turbulence) [11], resulting in difficult cleaning and an imbalance in the gas concentration in the cell linked to the measuring sensors. This is just one final thought on the simulation. The cause of this effect is the connection between tube and the spherical form is 90-degree.

As a result, we have created a new modification that uses a tangential arc to eliminate the corner on both sides (Figure 11). We maintain the same settings and setup as in Figure 10, the distribution of the airflow in Figure 11 is generally uniform and devoid of turbulence. In comparison to Figure 10, both corners, which are challenging to clean, are replaced by a tangential arc, which is simple to clean. Additionally, the air pressure inside the sensor array is raised, which is good for the way the sensors and the gas under measurement interact.

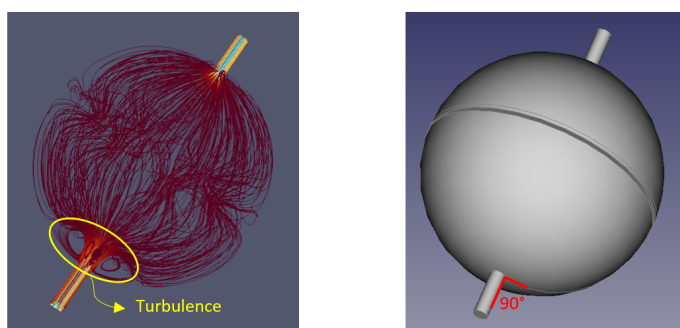


Figure 10: outcomes of fluid characteristic simulation (including the sensor array inside the chamber).

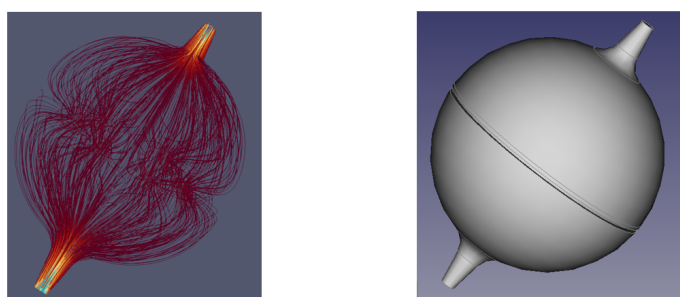


Figure 11: Simulation outcomes with the same flow rate and changed shape as in Figure 9.

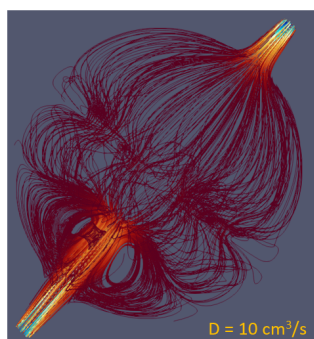


Figure 12: Simulation outcomes at 10cm<sup>3</sup>/s flow rate.

Figure 12, which keeps the same setup as Figure 11 but changes the flow rate parameter to 10cm<sup>3</sup>/s, displays the simulation results. As depicted in Figure 10, we may draw the conclusion that the system is constrained in terms of specific parameters, such as flow rate, and that it cannot be utilized for sensing measurements since the response will be non-confidential when employed (there will be a lot of turbulence). The values used for Figure 11 therefore provide the highest geometrical performance.

#### 4.2. Experience description

According to a huge number of tests carried out at the validation of built system dedicated to the quality control of agricultural products. This part results from tests done in order to come out with a significant percentage of system validation regarding data acquisition.

**Sample:** to ensure that the system works well (Figure 8) and with the lack of some sensor on the market, we worked with the most available sensors such as the MQ2 and MQ5 sensors, which are used for smoke detection and flammable gauges (for example butane), in this case our sample will be a lighter that can be contained in the cell, with a mechanical action so as to pull out the gas towards the outside.

**Technical part:** The sensor array consists of Six MOS chemiresistive type sensors (GS1, GS2, GS3, GS4, GS5 and GS6) capable of detecting various odors. To test the homogeneity of the chamber, the sensor array integrated with an Arduino UNO microcontroller having an on-chip analog to digital converter was used for data acquisition and plot it in the screen, two electro-valves are used for starting or stopping the gas flow coming from the pump.

A chemiresistive sensor or gas detector, principle of functionality in it [22], look like a resistor vary with different gases.

#### 4.3. Testing the sensors responses inside the chamber

We aim to test the response of these sensors (MQ2 and MQ5), based on the above procedure of the experiment, as shown in the Figure 13.

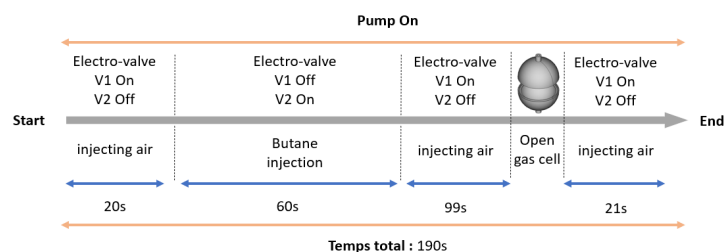


Figure 13: Procedure of the experiment.

The first 20 seconds are considered an initialization step for the sensors by airflow. Then we observe that there is a significant growth detected. This deduces that a new gas was captured. Finally, we return to the initial state performed another time by the injection of air, which will create a well on a signal decrease (the

gas cell fills with air) slowly. With this much-delayed decrease, we gave it a helping hand at time 178 s by manually opening the sensor chamber to release the butane gas in a faster way.

The sensor response is illustrated in the figure below:

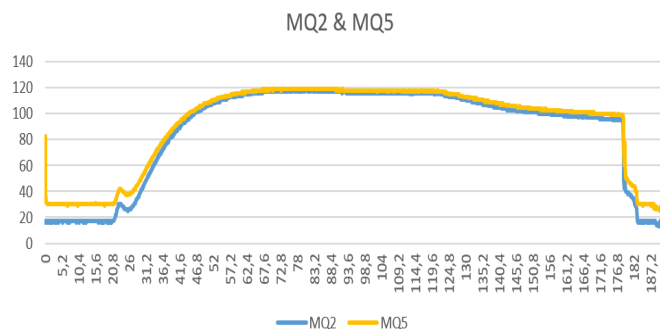


Figure 14: response of the butane gas flow from the sample cell to the measuring cell (C1 to C2 see Figure 9) observed by the sensors MQ2 and MQ5.

We can see the reaction of these sensors with butane gas, which we have an increasing after 20 seconds of butane flowing as shown in Figure 13, the amplitude gone from 20 as minimal to 120 analog values, and decreasing when we inject the air flow, with these responses we can see that the two sensors inside the chamber are functional, for reason to make sure that the system is working.

#### 4.4. Homogeneity test in the gas chamber

To evaluate the homogeneity of the gas sensor cell, the sensors must have almost the same answer, for this, the idea is that we will follow it is to work with only one sensor of type MQ2 at a time.

Since the electronic board supports 6 gas sensor locations, this involves 6 tests for each position, because only one measurement sensor was used. The following figure shows the procedure of the experiment:

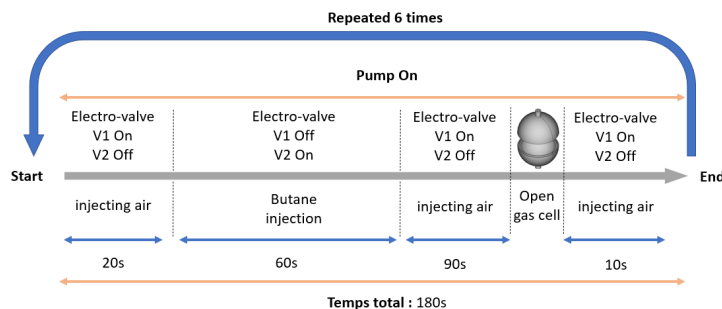


Figure 15: Homogeneity experiment procedure.

We respect as much as possible all the details of the experiment with the same condition as possible, the result is shown in the following figure.

By respecting the same conditions at each turn of the experiment, namely the concentration of gas in the sample cell in C1, the

opening and closing of the solenoid valves and the practical operation of the MQ2 sensor.

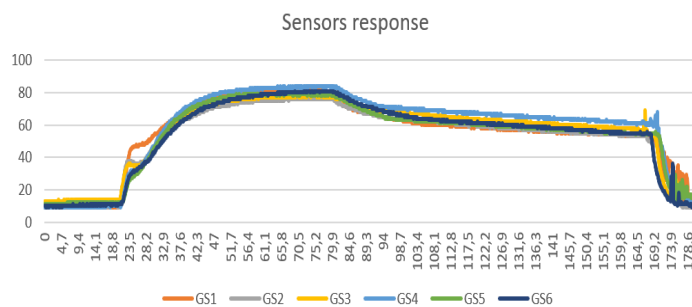


Figure 16: Homogeneity test response on the 6 sensor positions in C2 cell.

From these measurements it is found that they are almost homogeneous with some small deference either at the beginning of the transitional regime and the end of the measurement, the first is due to the fact that the concentration of gas mixture (air + butane) are not the same when it touches the sensor membrane (the sensitive part of the sensor against the gauze) for the 6 different positions on the electronic board, for the second is generated by the opening of the chamber causes a vibration of the sensor giving an observable disturbance at the measurement level.

But in general, these answers present a quality of measurement in the permanent diet that is most important in the data processing phase.

## 5. Conclusion and future work

For several uses in machine learning, including quality control, the e-nose system was an essential tool. The sensor chamber is crucial in the detection phase of any electronic nasal application. This investigation's analytical, which mainly focused on geometrical restrictions of e-nose design, discovered that, the spherical shape shows the highest geometrical performance for the debit due to several factors including volume and shape.

On the one hand, this practical work is an image that must respond to the simulation of the gas cell on the homogeneity of data capture according to the sensors. In the same way, we found a compatibility relationship between them, since the 6 sensors positions have an important reliability of the information, which gives the next step towards the processing and the pattern recognition phases.

## References

- [1] H. J. Ko and T. H. Park, "Bioelectronic nose and its application to smell visualization," *J. Biol. Eng.*, **10**(1), 17, Dec. 2016, doi: 10.1186/s13036-016-0041-4.
- [2] G. Ohloff., "Chemistry of odor stimuli.Experientia," **42**, 271-279., 1986.
- [3] S. S. Schiffman and T. C. Pearce, "Introduction to Olfaction: Perception, Anatomy, Physiology, and Molecular Biology," in *Handbook of Machine Olfaction*, T. C. Pearce, S. S. Schiffman, H. T. Nagle, and J. W. Gardner, Eds. Weinheim, FRG: Wiley-VCH Verlag GmbH & Co. KGaA, 2002, 1–31. doi: 10.1002/3527601597.ch1.
- [4] M. Harmouzi, A. Amari, and L. Masmoudi, "The conception of gas chamber for electronic nose," in *2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*,

- Meknes, Morocco, Mar. 2022, 1–4. doi: 10.1109/IRASET52964.2022.9737802.
- [5] N. S. Samiyani and M. M. Addi, “Characterization of Sensing Chamber Design for E-Nose Applications,” *J. Telecommun. Electron. Comput. Eng. JTEC*, **9**, 3–9, Art. no. 3–9, Dec. 2017.
- [6] P. Boeker, “On ‘Electronic Nose’ methodology,” *Sens. Actuators B Chem.*, vol. 204, 2–17, Dec. 2014, doi: 10.1016/j.snb.2014.07.087.
- [7] S. Kiani, S. Minaei, and M. Ghasemi-Varnamkhasti, “Instrumental approaches and innovative systems for saffron quality assessment,” *J. Food Eng.*, **216**, 1–10, Jan. 2018, doi: 10.1016/j.jfoodeng.2017.06.022.
- [8] E. L. Hines, P. Boilot, J. W. Gardner, and M. A. Gongora, “Pattern Analysis for Electronic Noses,” in *Handbook of Machine Olfaction*, T. C. Pearce, S. S. Schiffman, H. T. Nagle, and J. W. Gardner, Eds. Weinheim, FRG: Wiley-VCH Verlag GmbH & Co. KGaA, 2002, 133–160. doi: 10.1002/3527601597.ch6.
- [9] R. Gutierrez-Osuna, H. T. Nagle, B. Kermani, and S. S. Schiffman, “Signal Conditioning and Preprocessing,” in *Handbook of Machine Olfaction*, T. C. Pearce, S. S. Schiffman, H. T. Nagle, and J. W. Gardner, Eds. Weinheim, FRG: Wiley-VCH Verlag GmbH & Co. KGaA, 2002, 105–132. doi: 10.1002/3527601597.ch5.
- [10] A. Koocheki and E. Milani, “Saffron adulteration,” in *Saffron*, Elsevier, 2020, 321–334. doi: 10.1016/B978-0-12-818638-1.00020-4.
- [11] J. Roussel, “Cours de mécanique des fluides – femto-physique.fr,” 68.
- [12] “Understanding Pressure Measurement.” <https://www.meddeviceonline.com/doc/understanding-pressure-measurement-0001> (accessed Oct. 19, 2021).
- [13] S.-M. Jafari, M. Z. Tsimidou, H. Rajabi, and A. Kyriakoudi, “Bioactive ingredients of saffron: extraction, analysis, applications,” in *Saffron*, Elsevier, 2020, 261–290. doi: 10.1016/B978-0-12-818638-1.00016-2.
- [14] K.-T. Tang, S.-W. Chiu, C.-H. Pan, H.-Y. Hsieh, Y.-S. Liang, and S.-C. Liu, “Development of a Portable Electronic Nose System for the Detection and Classification of Fruity Odors,” *Sensors*, **10**(10), 9179–9193, Oct. 2010, doi: 10.3390/s101009179.
- [15] S. Kiani, S. Minaei, and M. Ghasemi-Varnamkhasti, “Integration of computer vision and electronic nose as non-destructive systems for saffron adulteration detection,” *Comput. Electron. Agric.*, **141**, 46–53, Sep. 2017, doi: 10.1016/j.compag.2017.06.018.
- [16] M. Ezhilan, N. Nesakumar, K. J. Babu, C. S. Srinandan, and J. B. B. Rayappan, “Freshness Assessment of Broccoli using Electronic Nose,” *Measurement*, **145**, 735–743, Oct. 2019, doi: 10.1016/j.measurement.2019.06.005.
- [17] K. Heidarbeigi, S. S. Mohtasebi, A. Foroughirad, M. Ghasemi-Varnamkhasti, S. Rafiee, and K. Rezaei, “Detection of Adulteration in Saffron Samples Using Electronic Nose,” *Int. J. Food Prop.*, **18**(7), 1391–1401, Jul. 2015, doi: 10.1080/10942912.2014.915850.
- [18] P. Le Maout, “Polyaniline nanocomposites based sensor array for breath ammonia analysis. Portable e-nose approach to non-invasive diagnosis of chronic kidney disease,” *Sens. Actuators B Chem.*, **274**, 616–626, Nov. 2018, doi: 10.1016/j.snb.2018.07.178.
- [19] L. Kumari, “Various techniques useful for determination of adulterants in valuable saffron: A review”, 3 March 2021, doi: <https://doi.org/10.1016/j.tifs.2021.02.061>
- [20] T. Hübert, “Chapter 12 - Electronic Noses for the Quality Control of Spices,” 115–124, 2016, doi: <https://doi.org/10.1016/B978-0-12-800243-8.00012-3>.
- [21] M. S. Hallé, M. L. Dufresne, and M. L. Lamarche, “MODÉLISATION DE LA QUALITÉ DE L’AIR DANS UNE UNITÉ DE BRONCHOSCOPIE : INFLUENCE DES STRATÉGIES DE VENTILATION,” 132.
- [22] M. A. Franco, P. P. Conti, R. S. Andre, and D. S. Correa, “A review on chemiresistive ZnO gas sensors,” *Sens. Actuators Rep.*, **4**, 100100, Nov. 2022, doi: 10.1016/j.snr.2022.100100.

## Prototype to Identify the Capacity in Cybersecurity Management for a Public Organization

Richard Romero Izurieta<sup>1</sup>, Segundo Moisés Toapanta Toapanta<sup>2,\*</sup>, Luis Jhony Caucha Morales<sup>3</sup>, María Mercedes Baño Hifong<sup>2</sup>, Eriannys Zharayth Gómez Díaz<sup>4</sup>, Oscar Marcelo Zambrano Vizueté<sup>4</sup>, Luis Enrique Mafla Gallegos<sup>5</sup>, José Antonio Orizaga Trejo<sup>6</sup>

<sup>1</sup> Faculty of Education Sciences; Universidad Estatal de Milagro (UNEMI), Milagro 091051, Ecuador

<sup>2</sup> Postgraduate Subsystems, Universidad Católica de Santiago de Guayaquil (UCSG), Guayaquil 090615, Ecuador

<sup>3</sup> Postgraduate School; Universidad Nacional de Tumbes, Tumbes 24001, Perú

<sup>4</sup> Research Department, Instituto Tecnológico Superior Rumiñahui, Sangolquí 171103, Ecuador

<sup>5</sup> Faculty of Engineering System, Escuela Politécnica Nacional (EPN), Quito 170525, Ecuador

<sup>6</sup> Information Systems Department (CUCEA), Universidad de Guadalajara, Guadalajara 44100, México

### ARTICLE INFO

Article history:

Received: 30 November, 2022

Accepted: 27 January, 2023

Online: 07 February, 2023

Keywords:

Cybersecurity

Management capacity

Public organizations

Security models

Security prototypes

### ABSTRACT

Public organizations are subjected to a complex security situation, which can be addressed by permanently strengthening and evaluating their cybersecurity capabilities. The objective of this research is to develop a model to identify the cybersecurity management capacity of public organizations. The deductive method was applied for the review and analysis of criteria, factors and variables related to cybersecurity capacity in public organizations. It resulted in a model to identify the Cybersecurity Management Capacity of public organizations, with its process to assess and categorize organizations according to their level of cybersecurity capacity. It was concluded that public organizations from developed countries in cybersecurity such as Spain have better capacities (greater than 60% CMC) than less developed countries such as Ecuador (less than 60% CMC), due to the cybersecurity context where these organizations operate. To obtain a high level of cybersecurity, public organizations must have the support of the governments of the different political divisions of a country, as well as permanent international collaboration in the field of cybersecurity.

## 1. Introduction

Security problems in public organizations in Ecuador are persistent; the authors propose a model based on strategic planning for the evaluation of information security [1].

Cyberattacks and the consequences suffered by organizations increased by 50% in 2021 [2]. Security and risk assessment tools are required to develop digital economies capable of coping with and recovering from challenging situations [3]. Cyber threats are now sophisticated and advanced, with greater impact and on a global scale, cyber risk has evolved and this implies that organizations and their capabilities to deal with these threats must also evolve; More than 4,000 ransomware attacks occur daily, with

financial losses of USD 265 billion, with an average system outage time of 19 days [4].

Given the complex security situation to which organizations are subjected, their capacities must be strengthened, with a holistic, proactive approach to prevention, permanently evaluating investments in security [5]. Organizations must work based on a well-articulated, shared strategic vision of IT, with a structure capable of ensuring improvements by making efficient use of available resources[6]. In addition to a strategic vision, controls must be implemented to ensure the information of the information and critical assets of the organization [7]. It is important that organizations have models, methodologies and tools to evaluate information security, to avoid suffering damages due to the intensification of sophisticated cyber-attacks [8].

Ecuador has a low level of capacity to combat cybercrime, related to a high rate of registered incidents according to statistics

\* Corresponding Author: Segundo Moisés Toapanta Toapanta, segundo.toapanta@cu.ucsg.edu.ec

from national and international organizations [9]. Ecuador Digital is the strategy to transform the country into an information and knowledge society, implementing digital government, the efficiency of public administration and digital adoption in the social and economic sectors, through three pillars: connectivity, efficiency and cybersecurity. and innovation and competitiveness [10].

The countries of Latin America are highly exposed to cyber attacks, due to their multiple deficiencies in the regulatory and institutional framework, infrastructure and other aspects, which is why they have a low level of cybersecurity capacity, although they have made efforts to improve these capabilities [11]. One of the significant advances for cybersecurity in Latin American countries is the Network of Cyber Incident Response Teams (CSIRTs) for the member states of the Organization of American States (OAS), in Ecuador it is called EcuCERT [12].

The European Union (EU) is one of the blocks with the greatest development of cybersecurity capabilities, it has defined strategies and objectives that member countries must meet, they are aware of the importance of the external context, both national and supranational, to strengthen their ability to cybersecurity [13].

The objective of this research is to develop a model to identify the cybersecurity management capacity of public organizations.

Why is it necessary to measure the cybersecurity capacity of public organizations?

It is necessary to identify the cybersecurity management capacity to know the current information security situation of public organizations, so that through a strategic IT perspective, the organization can constantly improve cybersecurity and maintain itself at an optimal level that allows preventing and mitigating risks. and cyber threats.

Considering the main factors of cybersecurity, used by organizations and states around the world, implies improving and adding capabilities that guarantee the Confidentiality, Integrity, and Availability of information and protect your critical IT assets.

The assessment of compliance with each of the criteria for each cybersecurity factor allows us to identify the capabilities that the organization has, which must be improved with a strategic vision.

In this process, the deductive method and exploratory research are used for the analysis of information related to cybersecurity capacity.

The main results obtained are: A management model for cybersecurity based on strategic planning; process and matrices for the evaluation of the Cybersecurity Management Capacity.

Public organizations from developed countries in cybersecurity such as Spain have better capacities (greater than 60% CMC) than less developed countries such as Ecuador (less than 60% CMC), due to the cybersecurity context where these organizations operate. Hence, to obtain a high level of cybersecurity, public organizations must have the support of the governments of the different political divisions of a country, as well as permanent international collaboration in the field of cybersecurity.

Managing cybersecurity optimally involves starting with strategic planning that allows directing the resources and

capabilities available to achieve the objectives established for the organization.

## 2. Materials and methods

### 2.1. Materials

The works that served as the basis for determining the main cybersecurity factors and variables in organizations are the following:

They define the cybersecurity culture, the main contributing factors and the metrics to evaluate organizations [14]. They evaluate the management of information security in public organizations [15]. To improve cybersecurity in public organizations, they recommend implementing a culture of Information Security [16]. They propose a conceptual model with a set of metrics to improve the efficiency of information security tasks [17]. Presents two models for the development of an information security assessment system for organizations [18]. Established a model of management success factors for information security in organizations [19]. They developed a security maturity model for organizations considering factors such as technology, people and infrastructure [20]. They recognize the key success factors of information security in organizations [21]. They analyzed how cybersecurity in organizations improves through the use of international standards and specific laws of a country [22]. They presented a conceptual model to manage the identity of the database of a public organization [23]. They present a prototype of a tool for security analysis and protection of organizations by joining component fault tree models and attack trees [24].

The levels of corruption in the local, national and international context are negatively related to the efficiency of investments in organizations. Some indicators that make it possible to measure corruption at the country level are the Corruption Perception Index (CPI) of Transparency International, the Corruption Control Index (CCI) of the World Bank and the Corruption Index of the International Country Risk Guide [25].

#### 2.1.1 Internal factors

In Table 1, we differentiate 4 internal factors, the first is the "Strategic" factor, which must start with the Strategic Planning of Information Security, in order to protect the public organization from cybersecurity risks and threats; In this way, the entire organization is aligned to the mission, vision and defined strategic objectives, which end in the execution of projects in each of the areas, considering all strategic, tactical and operational organizational levels. Within any strategic management it is important to know the current state of cybersecurity of the organization, to know what we must improve, always with the support of senior management, creating an organizational culture of security that includes all staff [26].

Table 1: Internal factors

Factor	Detail	Reference
Strategic	Safety culture and awareness, align senior management, management	[14,16,19,21]

	support, security policy, training and awareness	
Technology, infrastructure and resources	Resources, hardware y software	[15,19]
Organization / Management	Procedure and organization, norms, international standards, best practices, controls	[15–17, 19, 20, 22, 23]
Continuous improvement	Continuous improvement, risk assessment, security measurement, auditing, security analysis and protection	[17,18,20,23,24]

We call the other group of variables "Technology, infrastructure and resources", which are essential to operate and implement any management or project in the organization, such as human, technological, material resources, among others, as well as the physical infrastructure, networks, etc.

We call the third group of variables in the "Organization / Management" factor, which are the different organizational structures that obey the strategic need, which allows managing and controlling, considering the "Technology, infrastructure and resources" factor; includes processes and procedures, considering international standards and best practices, such as ISO/27001, ITIL, COSO, etc.

We call the fourth group of variables the "Continuous Improvement" factor; which is the implementation of a permanent management system, which ensures the control and monitoring of the operation of the controls and procedures carried out, as well as the constant improvement of what is working incorrectly to achieve the protection desired by the organization.

2.1.2. External factors

Public organizations are not isolated entities, they carry out their operations within a context that will affect their security[27]. The laws, state policies and other actions to curb cybercrime in each community, city or country, together with international cooperation, can positively or negatively affect the cybersecurity of an organization[15]. There is evidence that links the development of a good cybersecurity strategy in a country and the effective use of public resources can improve the cybersecurity of organizations[28]. We have called these external factors that affect the cybersecurity of public organizations: local, national and international context.

An effective cyber security approach must involve all levels of government, according to the political division of each country; Cyberspace is constantly evolving, as are attacks, threats and risks, which is why governments need to build resilient cybersecurity at all levels, so as not to be an easy target[29].

Cybersecurity Capacity Maturity Model for Nations (CMM), developed by the Global Cybersecurity Capacity Center (GCSCC),

at the University of Oxford, uses 5 dimensions: "Cybersecurity Policy and Strategy", "Cyber Culture and Society", "Education, Training and Skills in Cybersecurity", "Legal and Regulatory Frameworks" and "Standards, Organizations and Technologies". According to the 2020 Cybersecurity report of the Organization of American States, for the countries of Latin America and the Caribbean, the average maturity level is low, between 1 and 2, out of 5 levels of the CMM[30].

The Global Cybersecurity Index (GCI), is an initiative of the International Telecommunication Union (ITU) of the United Nations (UN), is based on 5 pillars: "legal measures", "technical measures", "organizational measures", "capacity development measures" and "cooperation measure"[31]. If we review the GCI ranking, the first 10 positions are: first place United States 100; second place United Kingdom and Saudi Arabia 99.54; third place Estonia 99.48; fourth place Korea, Singapore and Spain 98.52; fifth league Russia, Arab Emirates and Malaysia 98.06; sixth place Lithuania 97.93; seventh place Japan 97.82; eighth place Canada 97.67; ninth place France 97.6; 10th place India 97.5. Europe is the continent with the best positioned countries, we have 6 in the top 10.

The National Cyber Security Index (NCSI) measures the Cybersecurity of countries considering 12 indicators: "Development of cybersecurity policies", "Analysis and information on cyber threats, Education and professional development", "Contribution to global cybersecurity", "Protection of digital services", "Protection of essential services", "Electronic identification and trust services", "Protection of personal data", "Response to cyber incidents", "Cyber crisis management", "Fight against cybercrime" and "Military cyber operations"[32]. In the NCSI ranking, the first 10 countries belong to Europe, led by Greece 96.10, Lithuania 93.51, Belgium 93.51, Estonia 93.51, Czech Republic 92.21, Germany 90.91, Romania 89.61, Portugal 89.61, Spain 88.31 and Poland 87.01, which shows the progress of the European Union on cybersecurity issues.

2.2. Methods

2.2.1. First phase

A search was made for the information available on official websites and scientific databases on factors and variables that public organizations have used to analyze cybersecurity. The most common security problems suffered by organizations and their limitations to face cyber attacks were reviewed. Then the factors were analyzed and categorized into two groups, external and internal, related to the cybersecurity of public organizations.

2.2.2 Second phase

A conceptual model was designed to allow the evaluation of cybersecurity management capacity (CMC), based on the main internal and external factors found in the first phase, with a strategic approach, considering that it is one of the main deficiencies in organizations.

In order to quantify the cybersecurity management capacity of an organization, a calculation process was defined and measurement scales were created for the 5 fundamental factors of the conceptual model designed, based on variables that we can value found in the scientific literature and the practice of public



organizations. The calculation of CMC of an organization will be determined by the average of the evaluation of internal and external factors, both have the same weight.

For the factor criteria assessment scale, a standard scale between 0-10 is considered to obtain more precise results[33].

2.2.3. Third phase

To validate the Cybersecurity Management Capacity model, organizations in two different cybersecurity contexts or levels were assessed, the first context in a country developed in cybersecurity, belonging to the European Union and Spain, because it is within the top 10 both in the GCI index and in the NCSI. For the second context to compare, we have a country that still does not achieve good levels of cybersecurity, belonging to Latin America, such as Ecuador. For each context, three typical public organizations are simulated, with high, medium and low levels of cybersecurity management capacity.

To assess the external factor of the local, national and international context, the GCI 2020 ranking is taken into account, with 194 participating countries, where Spain is in position 4 with 98.52% and Ecuador is in position 119 with 26.30%[31]. To assess the level of corruption criterion, we use the Corruption Perception Index (CPI) 2021, with 180 participating countries, where Spain is ranked 34 with 61 points and Ecuador is ranked 105 with 35 points[34].

3. Results

The following results were obtained:

3.1. Cybersecurity management capacity model

Figure 1 shows the cybersecurity management capacity model (CMC), which groups 5 important factors found in the literature, 4 internal factors related to the capacities of the public organization and 1 external factor, related to the environment or context in which that the organization develops, which depending on the country can be more or less levels, depending on the respective political organization. This model proposes through the first "Strategic" factor, a perspective of IT strategic planning, which contemplates cybersecurity, to direct the organization to the achievement of its proposed objectives; For this, it must be based on the factors "Technology, infrastructure and resources", "Organization and Management", "Continuous Improvement" and "Local, National and International Context".

The mathematical model to calculate the % CMC will be given as follows:

$$F_i = \frac{\sum_{j=1}^n c_j}{n} \tag{1}$$

$$\%CMC = 1.25F_1 + 1.25F_2 + 1.25F_3 + 1.25F_4 + 5F_5 \tag{2}$$

where:

% CMC, is the measurement of the Cybersecurity Management Capacity of the public organization. With this % an organization can be categorized using Table 4.

F<sub>i</sub> is the assessment of each factor from i=1 to 5, according to Table 2. Each factor F is calculated by means of the average of the assessment of its criteria.

c<sub>j</sub> is the evaluation of the criteria with a score from 0 to 10, according to Table 3. For each factor F there can be criteria from j=1 to n.

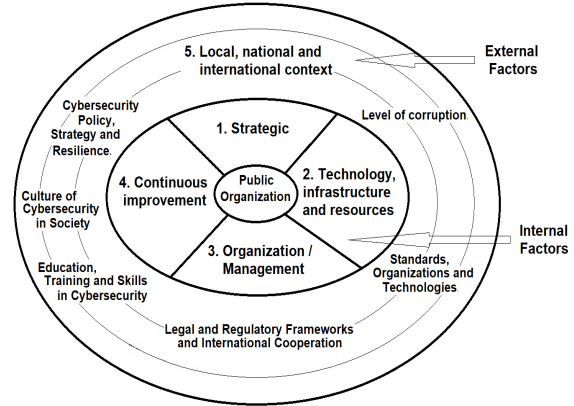


Figure 1: Cybersecurity Management Capacity Model

3.2. Process to quantify the CMC

In order to apply the CMC conceptual model and quantify the measurement, for each factor we established criteria to be evaluated for a public organization. Table 2 shows these factors with their respective criteria, which have been improved accordingly [1]. For the internal factors, practical criteria used in public organizations were considered, for the external factor we considered the 5 dimensions of the CMM model and a criterion of corruption levels was added[25]. We also defined the assessment scale for each factor criterion, which can be seen in Table 3; This scale starts from 0 to 3, which meets little or nothing, until reaching 9 to 10, where it meets all of that criterion, which is the maximum score that a one-factor criterion can have.

Table 2: Evaluation factors and considerations

Factor	Evaluation criteria
1. Strategic	1. Strategic IT planning. 2. Support from senior management. 3. Organizational culture of Safety. 4. Projects and action plans at the strategic, tactical and operational levels.
2. Technology, infrastructure and resources	1. Appropriate systems and technology. 2. Adequate IT infrastructure. 3. Sufficient human, financial, material and technological resources.
3. Organization / Management	1. Efficient and flexible organizational structure. 2. Control and management of all critical IT processes and assets. 3. International norms and standards, best IT Practices. 4. Control and management of IT strategic planning projects.
4. Continuous improvement	1. Incident and nonconformity management system.

	<ol style="list-style-type: none"> <li>Control and monitoring of incidents and nonconformities.</li> <li>Strategic planning considers reported incidents and nonconformities.</li> </ol>
5. Local, national and international context	<ol style="list-style-type: none"> <li>Cybersecurity Policy, Strategy and Resilience.</li> <li>Culture of Cybersecurity in Society.</li> <li>Education, Training and Skills in Cybersecurity.</li> <li>Legal and Regulatory Frameworks and International Cooperation.</li> <li>Standards, Organizations and Technologies.</li> <li>Level of corruption.</li> </ol>

Once all the criteria have been assessed, the average of each factor is calculated, to then determine the final weighted average of the 5 factors. In Table 4 we can see the 5 levels of the CMC model that a public organization can be categorized; starting from the "Initial", "Formative", "Administered", "Strategic" level, and ends with the highest level "Optimized", which should be the cybersecurity objective that a public organization must achieve.

Table 3: Factor Criteria Rating Scale

Scale	Value	Valuation Criterion
Very high	(9 – 10]	Meets all
High	(7 – 9]	Meets most
Medium	(5 – 7]	Partially complies
Low	(3 – 5]	Fulfills something
Very low	[0 – 3]	Little or no compliance

Table 4: CMC rating scale (In Organizations)

Scale	Value range	Assessment
optimized	(80 - 100]	Prepared
strategic	(60 - 80]	Consenting
managed	(40 - 60]	Vulnerable
formative	(20 - 40]	Danger
Initial	[0 - 20]	Helpless

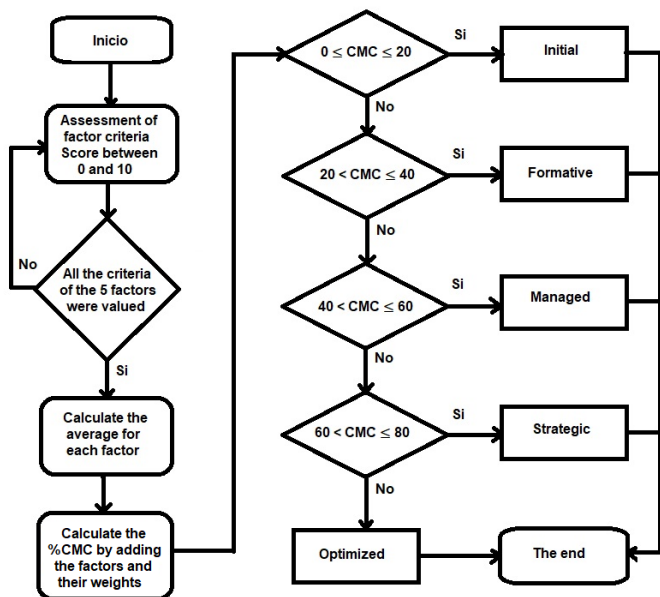


Figure 2: Process to calculate the Cybersecurity Management Capacity

Table 4 shows the CMC assessment scale, which categorizes an organization into 5 levels of capacity, similar to the CMM model; each level corresponds to 20% and goes from the Initial level that has a low assessment of the 5 factors, therefore, it is a defenseless organization, prone to attacks. The last level, on the other hand, speaks of an organization with a high rating in the 5 factors, which is prepared to prevent and combat any cyber-attack. Figure 2 shows the process for calculating the CMC; It consists of 4 stages: the assessment of the criteria for each factor considering Table 2 and 3, calculation of the average of the criteria assessments for each factor, calculation of the final average based on each factor and categorization of the organization according to the CMC of according to Table 4.

### 3.3. Validation of the CMC model

The validation of the CMC model of 2 different contexts Spain and Ecuador was carried out:

#### 3.3.1 Simulation of public organizations in Spain

Table 5 shows the final averages of each of the 5 factors for the simulation of 3 public organizations with High, Medium and Low levels of internal factors of the CMC model for the context of Spain. In Fig. 3 the results of the 3 organizations in Spain, where the red color represents the lack of CMC.

#### 3.3.2. Simulation of public organizations in Ecuador

Table 6 shows the final averages of each of the 5 factors for the simulation of 3 public organizations with High, Medium and Low levels of internal factors of the CMC model for the Ecuadorian context. In Fig. 4 the results of the 3 organizations from Ecuador, where the red color represents the lack of CMC.

Table 5: Spain context simulation

Factor	High	Medium	Low
1. Strategic	8.90	6.55	2.25
2. Technology, infrastructure and resources	9.50	5.68	3.14
3. Organization / Management	9.20	6.79	2.88
4. Continuous improvement	8.80	6.23	2.67
5. Local, national and international context	9.23	9.23	9.23
Final percentage	91.63%	77.69%	59.80%
Category	Optimized	Strategic	Managed

Table 6: Ecuador context simulation

Factor	High	Medium	Low
1. Strategic	8.90	6.55	2.25
2. Technology, infrastructure and resources	9.50	5.68	3.14
3. Organization / Management	9.20	6.79	2.88
4. Continuous improvement	8.80	6.23	2.67
5. Local, national and international context	2.78	2.78	2.78
Final percentage	59.38%	45.44%	27.55%
Category	Managed	Managed	Formative

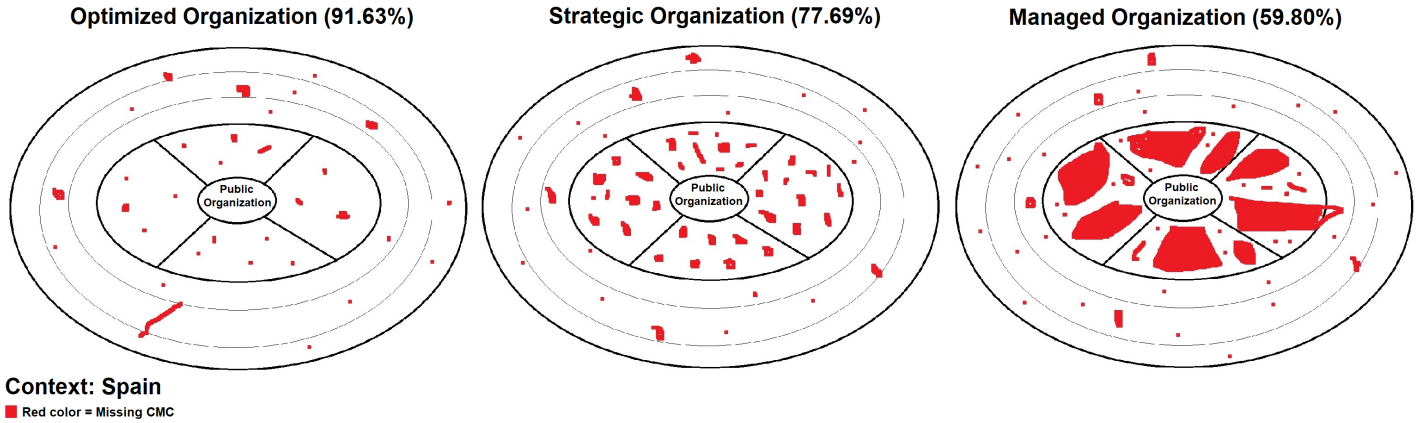


Figure 3: Simulation of public organizations in the context of Spain

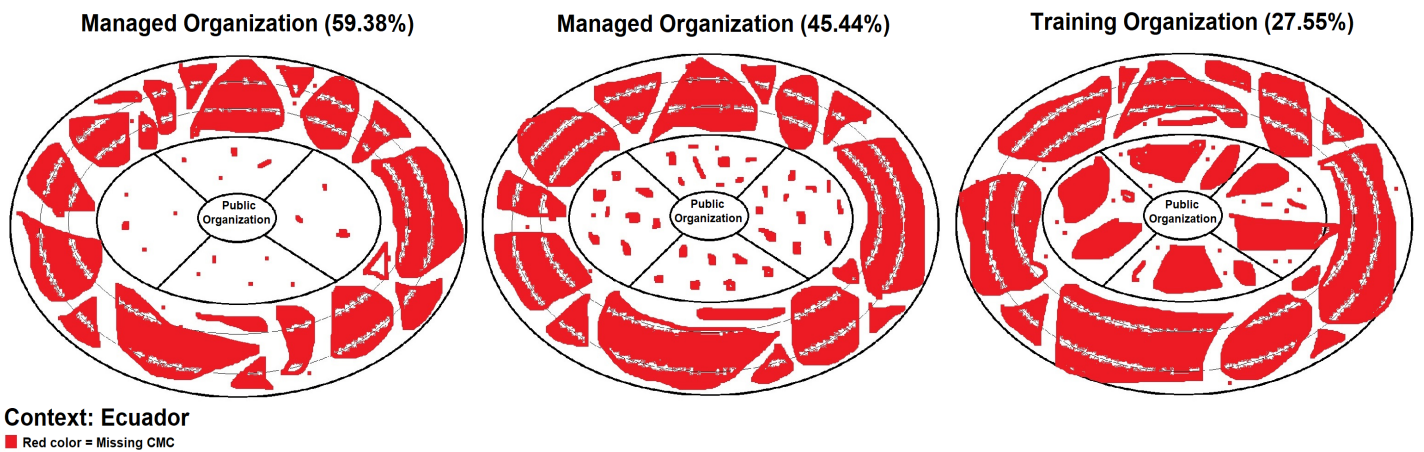


Figure 4: Simulation of public organizations in the Ecuadorian context

Fig. 3 shows the result of the simulation of the 3 organizations in the context of Spain, according to Table 5, the first with a valuation of high internal factors (Optimized), the second with medium factors (Strategic) and the third with factors low (Managed). For each organization, the graph resembles an onion because it has several layers, in the heart or center is the public organization, which is protected by the layer of internal factors of the CMC model, then there are several layers that belong to external factors, defined as a local, national and international context, the number of layers will depend on the political division of each country where the organization is located. We can see that the external factor of Spain greatly supports organizations in their cybersecurity management capacity, the parts marked in red represent the lack of capacity for each factor, which for the context of Spain are few.

Fig. 4 shows the result of the simulation of the 3 organizations in the context of Ecuador, according to Table 6, the first with a valuation of high internal factors (Administered), the second with medium factors (Administered) and the third with factors low (Formative). We can see that the external factor of Ecuador does little to support organizations in their cybersecurity management

capacity, the parts marked in red represent the lack of capacity for each factor and for this context in Ecuador there are many.

## 6. Discussion

A conceptual model is presented to determine the Cybersecurity Management Capacities in public organizations based on the most important factors found in the literature, both internal and external, where both groups of factors have the same weight. The 4 internal factors, "Strategic", "Technology, infrastructure and resources", "Organization / Management" and "Continuous improvement", form the first cybersecurity protective shield for the organization. The external factor "Local, national and international context" forms the following cybersecurity protective shield, which must work closely related to the internal factors of the organization, to achieve an "Optimized" category, which defines that the organization is prepared to face the possible computer attacks.

We can see in Figures 3 and 4 the simulation of organizations in 2 different contexts, such as Spain with excellent cybersecurity capabilities and Ecuador with limited cybersecurity capabilities. The results of the model show that, despite having similar internal factors in both contexts, the external factor makes the CMC superior for organizations in Spain, categorized as "Optimized" and "Strategic", while for Ecuador the CMC shows organizations

with problems in cybersecurity capacity, which can only be categorized as "Managed" and "Formative". This implies that in order to achieve optimal cybersecurity management capabilities, organizations have to strengthen not only internal factors, but also the external factor, which is the context in which the organization operates. The CMC model can be an important tool to know the current state of cybersecurity management capacity of organizations and to carry out periodic analyzes of the progress made to improve cybersecurity.

The proposed CMC model was developed from the perspective of assessing capabilities, based on the information that we know with certainty and have available, both internally and externally, from the context where the organization operates, which are actions of the different levels of government of a country and the international community. A large number of works reviewed in the literature maintain the perspective of evaluating cyber risk, based on unknown information, using probabilities of possible incidents and their effects; historical data is generally not available and subjective methods end up being used, such as expert judgment[35].

The results of the simulation of the CMC model for the public organizations showed notable differences for the compared contexts; In an advanced cybersecurity context like that of Spain, the vast majority of organizations will have a %CMC greater than 60%, which means that they may have a better chance of anticipating and resisting cyber-attacks. On the other hand, in the context of Ecuador, the vast majority of organizations will have a %CMC lower than 60%, which means that they are vulnerable, have many limitations of all their factors and are less likely to foresee and resist cyber-attacks. It is important to clarify that having a 100% CMC does not mean that the organization is safe from receiving cyber-attacks, it means that it has its cyber management capabilities developed to the maximum, in such a way that it can prevent or receive a minor impact, in such a way that business continuity is not threatened.

The ObservaCiber 2021 report shows results of important advances in cybersecurity of organizations in Spain, which supports the results of this work, with high %CMC found in organizations for the context of Spain[36]. Research carried out in Ecuador showed low levels of cybersecurity in public organizations[22]. International research also shows similar results, showing organizations with many cybersecurity problems, suggesting actions with government support and international collaboration to improve the low levels of cybersecurity shown[15].

## 7. Future work and conclusions

### 7.1. Future work

As future work, a validation of the model should be carried out in organizations of different contexts worldwide, in this way the CMC model can be perfected, validating the factors and criteria exhaustively, to ratify or modify them.

### 7.2. Conclusions

The Cybersecurity Management Capacities model allows to evaluate the current situation of the organization, to go gradually according to its resources and needs, to improve the CMC until

reaching an "Optimized" level, so that the organization has capacities that allow to foresee and protect your critical assets and sensitive information.

This CMC model highlights the need for public organizations to have the support of the governments of the different political divisions of a country, as well as permanent international collaboration in the field of cybersecurity. This is evidenced by the simulation carried out, where organizations from developed countries such as Spain have better capacities (greater than 60% CMC) than less developed countries such as Ecuador (less than 60% CMC), due to the cybersecurity context where these organizations operate.

Having a high % of CMC means that the organization has developed the necessary capabilities to be proactive and reactive in the face of possible attacks and cybersecurity problems, ensuring the continuity of the organization's operations and the reliability, integrity and availability of information. The CMC is a double protective shield, one internal and one external, but that does not imply that having a 100% CMC does not receive attacks and cyber threats, because that does not depend on the CMC.

## Conflict of Interest

The authors declare no conflict of interest.

## Acknowledgment

The authors thank the Doctorado en Estadística y Matemática Aplicada de la Universidad Nacional de Tumbes, Universidad Católica de Santiago de Guayaquil (UCSG), Instituto Tecnológico Superior Rumiñahui (ISTER), Escuela Politécnica Nacional (EPN), CUCEA Universidad de Guadalajara de México (UDG) and Secretaría de Educación. Superior, Ciencia, Tecnología e Innovación" (Senescyt).

## References

- [1] R. Romero I., L.J. Caucha M., S.M. Toapanta T., L.E. Mafla G., J.A. Orizaga T., "Analysis of the Information Security of Public Organizations in Ecuador," in The 2021 International Conference on Computational Science and Computational Intelligence, IEEE: 823–829, 2021, doi:10.1109/CSCI54926.2021.00195.
- [2] Check-Point-Research, Check Point Software's 2022 Security Report: Global Cyber Pandemic's Magnitude Revealed, 2022.
- [3] World Economic Forum, Global Cybersecurity Outlook 2022, Cologny/Geneva.
- [4] U.L.C.S. Andrew Morrison, Principal, Cyber Security Landscape 2022.
- [5] Check-Point-Research, CYBER SECURITY REPORT, 2021.
- [6] G.M. Jonathan, K.S. Hailemariam, B.K. Gebremeskel, S.D. Yalaw, "Public Sector Digital Transformation: Challenges for Information Technology Leaders," in 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 1027–1033, 2021, doi:10.1109/IEMCON53756.2021.9623161.
- [7] W.A. Conklin, D. Shoemaker, "Cyber-resilience: Seven steps for institutional survival," *EDPACS*, **55**(2), 14 – 22, 2017, doi:10.1080/07366981.2017.1289026.
- [8] M. Nassar, J. Khoury, A. Erradi, E. Bou-Harb, "Game Theoretical Model for Cybersecurity Risk Assessment of Industrial Control Systems," in 2021 11th IFIP International Conference on New Technologies, Mobility and Security, NTMS 2021, 2021, doi:10.1109/NTMS49979.2021.9432668.
- [9] M. Ron, W. Fuertes, M. Bonilla, T. Toulkeridis, J. Díaz, "Cybercrime in Ecuador, an exploration, which allows to define national cybersecurity policies," in 2018 13th Iberian Conference on Information Systems and Technologies (CISTI), 1–7, 2018, doi:10.23919/CISTI.2018.8399357.
- [10] MINTEL, Libro Blanco de Territorios Digitales en Ecuador, Ministerio de

- Telecomunicaciones y de la Sociedad de la Información, Quito, 2019.
- [11] A. Urbanovics, "Cybersecurity Policy-Related Developments in Latin America," *Academic and Applied Research in Military and Public Management Science*, **21**(1), 79–94, 2022.
- [12] S. Creese, W.H. Dutton, P. Esteve-González, "The social and cultural shaping of cybersecurity capacity building: a comparative study of nations and regions," *Personal and Ubiquitous Computing*, **25**(5), 941–955, 2021, doi:10.1007/s00779-021-01569-6.
- [13] S. Creese, W.H. Dutton, P. Esteve-Gonzalez, M. Goldsmith, E. Nagyfejeo, J. Saunders, B. von Solms, C. Weisser Harris, "The Solution is in the Details: Building Cybersecurity Capacity in Europe," Available at SSRN 4178109, 2022.
- [14] B. Uchendu, J.R.C. Nurse, M. Bada, S. Furnell, "Developing a cyber security culture: Current practices and future needs," *Computers and Security*, **109**, 2021, doi:10.1016/j.cose.2021.102387.
- [15] E.K. Szczepaniuk, H. Szczepaniuk, T. Rokicki, B. Klepacki, "Information security assessment in public administration," *Computers & Security*, **90**, 101709, 2020, doi:10.1016/j.cose.2019.101709.
- [16] A. Nasir, R.A. Arshah, M.R. Ab Hamid, S. Fahmy, M.A. Bakar, "Information security culture model for malaysian organizations: A review," *International Journal*, **9**(1.3), 117–121, 2020, doi:10.30534/ijatcse/2020/1691.32020.
- [17] F.Ö. Sönmez, "A conceptual model for a metric based framework for the monitoring of information security tasks' efficiency," in *Procedia Computer Science*, 181–188, 2019, doi:10.1016/j.procs.2019.09.459.
- [18] R. Hoffmann, J. Napiórkowski, T. Protasowicki, J. Stanik, "Measurement models of information security based on the principles and practices for risk-based approach," *Procedia Manufacturing*, **44**, 647–654, 2020, doi:10.1016/j.promfg.2020.02.244.
- [19] R. Diesch, M. Pfaff, H. Kremer, "A comprehensive model of information security factors for decision-makers," *Computers & Security*, **92**, 101747, 2020, doi:10.1016/j.cose.2020.101747.
- [20] O.M.M. Al-Matari, I.M.A. Helal, S.A. Mazen, S. Elhennawy, "Adopting security maturity model to the organizations' capability model," *Egyptian Informatics Journal*, **22**(2), 193–199, 2021, doi:10.1016/j.eij.2020.08.001.
- [21] K. Arbanas, N. Žajdela Hrustek, "Key success factors of information systems security," *Journal of Information and Organizational Sciences*, **43**(2), 131–144, 2019, doi:10.31341/jios.43.2.1.
- [22] S.M. Toapanta, A. Jimenez, L.E. Mafla, "An approach of national and international cybersecurity laws and standards to mitigate information risks in public organizations of Ecuador," in *Proceedings of the 2019 2nd International Conference on Education Technology Management*, 61–66, 2019, doi:10.1145/3375900.3375909.
- [23] M. Toapanta, E. Mafla, J. Orizaga, "Conceptual model for identity management to mitigate the database security of the registry civil of Ecuador," *Materials Today: Proceedings*, **5**(1, Part 1), 636–641, 2018, doi:https://doi.org/10.1016/j.matpr.2017.11.127.
- [24] B. Kruck, P. Munk, D. Angermeier, "Safe and Secure: Mutually Supporting Safety and Security Analyses with Model-Based Suggestions," in *Proceedings - 2021 IEEE International Symposium on Software Reliability Engineering Workshops, ISSREW 2021*, 172 – 181, 2021, doi:10.1109/ISSREW53611.2021.00061.
- [25] X.M. Nguyen, Q.T. Tran, "Corruption and corporate investment efficiency around the world," *European Journal of Management and Business Economics*, **31**(4), 425 – 438, 2022, doi:10.1108/EJMBE-11-2020-0321.
- [26] H. Serna Gómez, others, *Gerencia estratégica. Planeación y gestión, teoría y metodología*, 3R Editores, 2008.
- [27] L. Masilela, D. Nel, "The role of data and information security governance in protecting public sector data and information assets in national government in South Africa," *Africa's Public Service Delivery and Performance Review*, **9**, 10, 2021, doi:https://doi.org/10.4102/apsdpr.v9i1.385.
- [28] Z. Li, X. Guo, Q. He, "A Study of Chinese Policy Attention on Cybersecurity," *IEEE Transactions on Engineering Management*, 2020, doi:10.1109/TEM.2020.3029019.
- [29] M. Masombuka, M. Grobler, P. Duvenage, "Cybersecurity and local Government: Imperative, Challenges and Priorities," in *ECCWS 2021 20th European Conference on Cyber Warfare and Security*, 285, 2021.
- [30] BID - OEA, *Ciberseguridad riesgos, avances y el camino a seguir en América Latina y El Caribe*, 2020.
- [31] G.C. Index, "URL: <https://www.itu.int/myitu/-/media/Publications/2021-Publications>," *Global-Cybersecurity-Index-2020*. Pdf [in English], 2020.
- [32] e-Governance Academy Foundation, *National Cyber Security Index NCSI, CSIRT*, 2020.
- [33] K.S. Crandall, "Risk Assessments: A Weighted Score Approach to Improving Risk Management Decisions," in *2020 Intermountain Engineering, Technology and Computing, IETC 2020*, 2020, doi:10.1109/IETC47856.2020.9249164.
- [34] Transparency international, *Corruption Perceptions Index CPI*, 2021.
- [35] M. Battaglioni, G. Rafeiani, F. Chiaraluca, M. Baldi, "MAGIC: A Method for Assessing Cyber Incidents Occurrence," *IEEE Access*, **10**, 73458 – 73473, 2022, doi:10.1109/ACCESS.2022.3189777.
- [36] ObservaCiber, *Indicadores sobre confianza digital y ciberseguridad en España y la Unión Europea*, 2021.

## Metaheuristic Optimization Algorithm Performance Comparison for Optimal Allocation of Static Synchronous Compensator

Abdulrasaq Jimoh<sup>1</sup>, Samson Oladayo Ayanlade<sup>\*2</sup>, Emmanuel Idowu Ogunwole<sup>3</sup>, Dolapo Eniola Owolabi<sup>4</sup>, Abdulsamad Bolakale Jimoh<sup>5</sup>, Fatina Mosunmola Aremu<sup>6</sup>

<sup>1</sup>Obafemi Awolowo University, Department of Electronic and Electrical Engineering, Ile-Ife, Nigeria

<sup>2</sup>Lead City University, Department of Electrical and Electronic Engineering, Ibadan, Nigeria

<sup>3</sup>Cape Peninsula University of Technology, Department of Electrical, Electronic and Computer Engineering, Cape Town, South Africa

<sup>4</sup>Ladoke Akintola University of Technology, Department of Electronic and Electrical Engineering, Ogbomoso, Nigeria

<sup>5</sup>University of Ilorin, Department of Electrical and Electronic Engineering, Ilorin, Nigeria

<sup>6</sup>Kwara State University, Department of Electrical Electronic Engineering, Malete, Nigeria

### ARTICLE INFO

Article history:

Received: 25 December, 2022

Accepted: 25 January, 2023

Online: 07 February, 2023

Keywords:

FACTS devices

STATCOM

Power loss

Voltage profile

Particle swarm optimization

Firefly algorithm

### ABSTRACT

The relevance of static synchronous compensator (STATCOM) controllers in controlling power network parameters is causing them to be included in contemporary networks. But for the intended objectives to be attained, the best device positioning and parameter settings are essential. This work compares the performance of the particle swarm optimization (PSO) and firefly algorithm (FA) in sizing and placing a STATCOM device for the dual objectives of loss reduction and voltage deviation abatement. The effective mitigation of network loss and voltage fluctuations in the network will be achieved by the deployment of the efficient method during device allocation. While PSO and FA were taken into consideration due to their computational efficiency among other metaheuristic algorithms, STATCOM was chosen from among the Flexible Alternating Current Transmission System (FACTS) controllers as a consequence of its reactive power compensation capability. The MATLAB software was used to implement the simulations on an IEEE 14-bus system. When STATCOM was optimized with PSO and FA, it resulted in active power loss reductions of 432 and 733 kW, respectively, and reactive power loss reductions of 1622 and 2100 kVAr, respectively. As a result, the reductions in voltage variation and power losses in this instance show some benefits of FA over PSO. Additionally, this work has shown that metaheuristic algorithms are beneficial for allocating FACTS devices.

## 1. Introduction

In current use, a power system is a system made up of a large number of power plants, transmission lines, loads, and transformers [1–2]. Increased power consumption causes transmission lines to become overloaded, which makes the power systems unstable. The system must thus operate very near its stability limit. This typically leads to a poor voltage profile and considerable network power loss [3–4].

The deployment of Flexible Alternating Current Transmission System (FACTS) devices and the building of new transmission lines are two options for addressing the problem of the power system overloading [5]. The construction of new power generation and the upgrading of transmission lines to reduce line congestion are both fraught with challenges. Increased load demands, constraints on the economy and the environment, and power networks operating nearer to their stability limits are all implications of the reorganization of the electrical sector [6]. For the aforementioned reasons, the power networks frequently encounter losses and voltage instability,

\*Corresponding Author: Samson Oladayo Ayanlade, Lead City University, +2348062786683, samson.ayanlade@lcu.edu.ng

[www.astesj.com](http://www.astesj.com)

<https://dx.doi.org/10.25046/aj080114>

which can result in voltage collapse. Sustaining the system's stability and safety is therefore a crucial and challenging problem.

To improve system stability and security, several strategies, including reactive power compensation (RPC) and phase shifting, are used [7]. The RPC is the strategy that is most frequently employed and well-liked among them since power networks are mostly reactive. Reactive power is required to maintain voltage magnitudes for transmitting active power across transmission lines. The primary source of power losses is the use of reactive power above the threshold set by the generators. By utilizing compensators, power losses may be reduced to a minimum. Different types of RPCs are employed in power networks to compensate for reactive power [8].

Power electronics and FACTS device advancements have made it possible to manage line flows, reduce overall system loss, and keep the voltage profile within permissible bounds in a power system [9]. FACTS are regulators that may alter several features of a transmission network. Through system parameter management, they also possess the capacity to swiftly and seamlessly consume or provide reactive power to the networks. These allow for voltage control on a specific bus.

Different categories have been established for FACTS devices. The work by [10] demonstrated the modeling of the FACTS device and its integration into power flow investigations. The position of the STATCOM, a shunt-type FACTS regulator, in the grid significantly affects losses and voltages and is primarily employed by power engineers for reactive power adjustment. The objective of STATCOM placement, an optimization issue, is to minimize power loss while respecting system constraints [11]. Power flow equations are utilized to demonstrate equality limitations, while upper and lower voltage limits are employed to represent inequality constraints. Swarm intelligence and population-based optimization techniques are frequently used to determine the ideal sizes for the devices, while load flow approaches continue to be a viable tool for determining the precise position for placement of these regulators.

FACTS allocation problems have been addressed using a variety of metaheuristic techniques, including Tabu Search (TS), Bat Algorithms (BAT), Whale Optimization Algorithm (WOA), Ant Lion Optimization (ALO) Algorithm, Simulated Annealing (SA), Artificial Bee Colony (ABC), etc. To boost network transfer performance, ABC was utilized by [12] to deploy FACTS regulators in the best possible way. To enhance the loadability of a power system, GA was utilized to efficiently deploy FACTS regulators in a power network. To minimize voltage magnitude changes and losses, BFOA was used by [13], [14] to determine the best location for UPFC devices. However, there has not been much research done to date to compare the effectiveness of these techniques in FACTS regulator optimization for transmission network capability improvement. Among all the metaheuristic optimization methods, the FA and PSO are two of the most efficacious. The FA was developed based on the distinctive ways that fireflies attract one another.

On the other hand, the PSO took inspiration from how insects behave while searching for food. Both the FA and PSO have been demonstrated to be reliable methods for resolving optimization problems, particularly in power systems. Thus, the efficacies of the FA and PSO in solving optimization problems in power systems cannot be overemphasized.

In this study, the STATCOM controller's allocation to enhance network voltage and diminish active and reactive losses is discussed. The implementations of PSO and FA for locating this regulator were described and applied to the IEEE 14-bus system because of their quick convergence and precision compared to other techniques. Two stages of the research were carried out: To begin with, a load flow study was done to find the buses that were over the typical range of permissible voltages. Second, PSO and FA methods were used for sizing the device needed for loss minimization. This study makes a contribution by contrasting the effectiveness of PSO and FA for deploying STATCOM controllers to enhance network functionality. Also, this study is novel in that it implements two separate metaheuristic optimization approaches to allocate STATCOM in the best way possible and determines which methodology is more effective; as a result, it assists power system engineers in society in adopting the quickest and most effective technique for resolving power system issues encountered in society to boost the general standard of living.

## 2. Model of STATCOM Controller

This controller is a regulator used for reducing transmission losses and alleviating voltage magnitude violation problems. It is made up of a parallel-connected controller and a static VAR generator, which uses different switching patterns within its converter to generate or absorb reactive power. To provide a sufficient supply of electricity, STATCOM corrects for reactive power in the electricity grids. When deployed, it moves more quickly between supplying and consuming reactive power, minimizing power losses and voltage fluctuations.

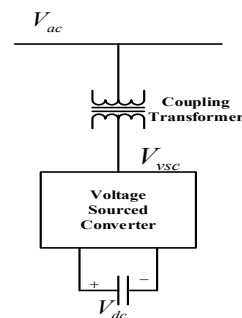


Figure 1: STATCOM controller configuration [15]

### 2.1. Mode of Operation

A simple STATCOM arrangement is shown in Fig. 1. It comprises a connecting transformer, a capacitor, and a voltage source converter (VSC). A series of three-phase voltages are created from the DC voltage by the VSC. The coupling transformer's functions include connecting the VSC to the high voltage side and preventing short circuits in the DC capacitor

[14]. A change in 3-phase converter voltage  $V_{vsc}$  varies the reactive supply to the network. If the STATCOM output voltage  $V_{vsc}$  more than the network's voltage  $V_{ac}$  (i.e.,  $V_{vsc} > V_{ac}$ ), the controller injects reactive power to the grid. Furthermore, if  $V_{vsc}$  does not exceed  $V_{ac}$  (i.e.,  $V_{vsc} < V_{ac}$ ), the STATCOM consumes reactive power from the grid. However, when  $V_{vsc}$  and  $V_{ac}$  the same (i.e.,  $V_{vsc} = V_{ac}$ ), the STATCOM is in standby mode.

## 2.2. STATCOM Power Flow Model

To control voltage, STATCOM either absorbs or provides reactive power to the network. The STATCOM connection at bus  $m$  is shown in Fig. 2.

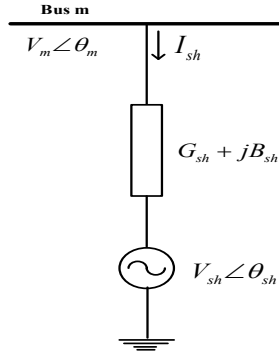


Figure 2: STATCOM Controller Equivalent

Bus  $m$  load flow equations following STATCOM deployment are stated as (1)–(4).

$$P_m = P_{sh} + \sum_{j=1}^N |V_m| |V_j| |Y_{mj}| \cos(\theta_{mj} - \delta_{mj}) \quad (1)$$

$$Q_m = Q_{sh} + \sum_{j=1}^N |V_m| |V_j| |Y_{mj}| \sin(\theta_{mj} - \delta_{mj}) \quad (2)$$

$$P_{sh} = G_{sh} |V_m|^2 - |V_m| |V_{sh}| |Y_{sh}| \cos(\theta_{msh} - \delta_{sh}) \quad (3)$$

$$Q_{sh} = B_{sh} |V_m|^2 - |V_m| |V_{sh}| |Y_{sh}| \sin(\theta_{msh} - \delta_{sh}) \quad (4)$$

where,  $V_m \angle \theta_m$ ,  $V_{sh} \angle \theta_{sh}$  = voltage at bus  $m$  and at STATCOM, respectively,  $P_m$ ,  $Q_m$  and  $P_{sh}$ ,  $Q_{sh}$  = bus  $m$  active and reactive power, and STATCOM,  $Y_{sh}$ ,  $G_{sh}$  and  $B_{sh}$  = STATCOM's admittance, conductance, and susceptance,  $Y_{mj} \angle \delta_{mj}$  = admittance of the line,  $N$  = number of buses.

## 3. Formulation of Problem

The optimum location of FACTS controllers to reduce losses is written as [16]:

Minimize  $f(x, \sigma)$

subject to

$$\begin{aligned} g(x, \sigma) &= 0 \\ h(x) &< 0 \\ x_l &< x < x_u \end{aligned} \quad (5)$$

where,  $g(x)$ ,  $h(x)$  = equality and inequality constraints,  $f(x)$  = total branch loss,  $\sigma$  = system load data,  $x_l$  and  $x_u$  = the minimum and maximum range.

The solution approach entails optimizing the objective function while satisfying the network restrictions, which include the load flow equations, voltage restrictions, and control parameter bounds [17].

### 3.1. Objective Function

This is done primarily to reduce overall active loss while remaining within the constraints [18].

$$\min \sum_{k \in N_g} P_{kloss} = \sum_{k \in N_g} g_k (V_i^2 + V_j^2 - 2V_i V_j \cos \theta_{ij}) \quad (6)$$

where,  $g_k$  = conductance in p.u.,  $k = (i, j)$ ,  $i \in N_B$  is the bus number,  $V_i$  and  $V_j$  = voltage magnitudes in p.u.,  $j \in N_i$  = bus number adjusted to bus  $i$ .

### 3.2. Equality Constraints

Each particle power flow equation is represented by (7)–(8). The load flow solution employs the Newton-Raphson approach.

$$P_{gi} - P_{Li} - V_i \sum_{j \in N} V_j (g_{ij} \cos \theta_{ij} + B_{ij} \sin \theta_{ij}) = 0 \quad (7)$$

$$Q_{gi} - Q_{Li} - V_i \sum_{j \in N} V_j (g_{ij} \sin \theta_{ij} + B_{ij} \cos \theta_{ij}) = 0 \quad (8)$$

where,  $B_{ij}$  = susceptance of the branch.

### 3.3. Inequality Constraints

The load and generator voltages, capacitive reactive power and transformer-tap settings, active and reactive line flow restriction, and power injection are all written as

$$V_i^{\min} \leq V_i \leq V_i^{\max}, \quad i \in N_B \quad (9)$$

$$Q_{gi}^{\min} \leq Q_{gi} \leq Q_{gi}^{\max}, \quad i \in N_g \quad (10)$$

$$Q_{ci}^{\min} \leq Q_{ci} \leq Q_{ci}^{\max} \quad (11)$$

$$T_k^{\min} \leq T_k \leq T_k^{\max} \quad (12)$$

$$S_l \leq S_l^{\max} \quad (13)$$

### 3.4. Fitness Function Formulation

It is written as

$$F_P = \sum_{q \in N} P_{qloss} + PF \quad (14)$$

The PF, which is the penalty function, is written as in (15).



$$q_1 \times \sum_{i=1}^{N_G} f(Q_{gi}) + q_2 \times \sum_{i=1}^N f(V_i) + q_3 \times \sum_{m=1}^{N_L} f(S_{lm}) \quad (15)$$

And  $q_1, q_2, q_3$  are penalty factors.

$$f(x) = \begin{cases} 0, & \text{if } x^{\min} \leq x \leq x^{\max} \\ (x - x^{\max})^2, & \text{if } x > x^{\max} \\ (x^{\min} - x)^2, & \text{if } x < x^{\min} \end{cases} \quad (16)$$

where,  $x^{\min}$  and  $x^{\max}$  = control parameters.

#### 4. Particle Swarm Optimization

PSO, an algorithm influenced by nature, was created in 1995 [19]. This algorithm uses particle populations to identify the optimum solution. Each particle is taken into account as a potential solution throughout the search process.

The phrases "particle," "swarm," "position," "swarm fitness," " $P_{best}$ ," " $g_{best}$ ," and the maximum and minimum permitted velocity values are all related to PSO.

Particles are generated at random by the method inside the scope of the function domain. The optimum position that individual particle  $i$  has found in the search space is shown by its current velocity ( $v$ ), personal best position ( $y_i$ ), and current position ( $x$ ). Every particle in a  $d$ -dimensional area tracks them according to:  $x_i = (x_{i1}, x_{i2}, \dots, x_{id})$ ,  $v_i = (v_{i1}, v_{i2}, \dots, v_{id})$ , and  $P_{best} = (P_{besti1}, P_{besti2}, \dots, P_{bestid})$ .

If there are  $s$  particles in the swarm.

Then,  $i \in I, \dots, s$ .

$$y_i(t+1) = \begin{cases} y_i(t) & \text{if } f(y_i(t) \leq f(x_i(t+1))) \\ x_i(t+1) & \text{if } f(y_i(t) > f(x_i(t+1))) \end{cases} \quad (17)$$

$$\begin{aligned} \hat{y}(t) &= \min \{f(y), f(\hat{y}(t))\} \\ y &\in \{y_0(t), y_1(t), \dots, y_s(t)\} \end{aligned} \quad (18)$$

At each iteration, (17) and (18) update each particle. For each dimension  $j \in 1 \dots n$ , if  $x_{ij}$ ,  $y_{ij}$ , and  $v_{ij}$  be the  $j^{\text{th}}$  dimension present position, personal best position and velocity of the  $i^{\text{th}}$  particle. The new velocity is given by (19).

$$v_{i,j}(t+1) = wv_{i,j}(t) + c_1r_{1,j}(t)[y_{i,j}(t) - x_{i,j}(t)] + c_2r_{2,j}(t)[\hat{y}_{i,j}(t) - x_{i,j}(t)] \quad (19)$$

To determine the particle's new position, the new velocity is added to its present position.

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (20)$$

To reduce the likelihood of the particle exiting the search space, all dimensional values of  $v_i$  are restricted to  $[-v_{max}, v_{max}]$ . The  $v_{max}$  is determined by (21).

$$v_{max} = k \times x_{max}, \quad \text{where } 0.1 \leq k \leq 1.0 \quad (21)$$

where,  $x_{max}$  = domain space search,  $c_1$  and  $c_2$  = coefficients of acceleration.

The PSO convergence behavior is controlled by the inertial weight, which is obtained using (22).

$$w = w_{max} - \frac{w_{max} - w_{min}}{itera_{max}} \cdot itera \quad (22)$$

where,  $itera_{max}$  = maximum number of iteration,  $itera$  = number of iteration,  $w_{max}$  and  $w_{min}$  = maximum and minimum weighting factor.

##### 4.1. PSO Implementation Algorithm for STATCOM Allocation

The IEEE 14-bus system was utilized to implement PSO. The particle placements were influenced by the initial control variable limitations. Computing the fitness value represented by (14), with the intention of reaching the reduced global best, yielded evaluations of the control variables. The steps for implementing the technique are as follows:

- The population size, total number of iterations, and all control parameters are specified.
- Set iteration number = 0.
- Create the populations and velocities of the particles.
- For loss calculations, run the Newton-Raphson power flow for each individual particle.
- Determine the fitness value for each particle by (14).
- Determine the  $P_{best}$  and  $g_{best}$  for each particle.
- Let  $iteration = iteration + 1$ .
- If there is a voltage restriction breach, the velocity and displacement of each individual particle are calculated using (19).
- Find the new location of each particle by (20).
- To calculate the power, run the Newton-Raphson power flow for each particle.
- Using (14), find the fitness value for each particle.
- If the particle's current fitness  $P$  is higher than  $P_{best}$ , set  $P_{best}$  to equal  $P$ .
- Set  $g_{best}$  to  $P_{best}$ .
- Up until the allotted iteration's number is reached, continue from step 7.

The smallest loss values from the relevant fitness value are used to calculate the parameters of  $g_{best}$  and the optimum values for the control parameters.

#### 5. Firefly Algorithm

Yang created this algorithm, which is a method for tackling challenging optimization issues quickly [20, 21].

##### 5.1. Firefly Behavior

According to the inverse-square law, the relationship between the intensity of light,  $I$ , and distance,  $r$ , is inverse. Due to this, the majority of fireflies may be seen at night for a brief period of time, such as a few hundred meters, which is sufficient for flies to converse. A potentially optimizable objective function is used to simulate the flashing light.

### 5.2. Implementation of Firefly Algorithm

There are three fundamental presumptions that should be taken into account and are stated below [22] for simplicity in the FA description:

- Fireflies have no gender.
- As the distance between fireflies grows, both attractiveness and brightness decrease.
- The objective function's terrain influences the firefly brightness.

The objective functions used in FA for the optimization problem are brightness and light intensity. Finding the optimal solution is similar to being drawn to and moving toward the firefly that is brighter [23].

### 5.3. Light Intensity and Attractiveness

FA is influenced by two variables: light intensity fluctuation and attractiveness formation.

The brightness of the firefly  $i$  and the distance between the two fireflies are both factors in the attractiveness,  $I$ , of the firefly  $i$  to the firefly  $j$  [24]. The expression for light's intensity, which changes with distance, is stated as (23).

$$I_{(r)} = \frac{I_s}{r^2} \quad (23)$$

where,  $I_{(r)}$  = light intensity,  $I_s$  = intensity of source.

The intensity is expressed as (24).

$$I_{(r)} = I_0 e^{-\gamma r} \quad (24)$$

To prevent singularity at  $r = 0$ , (23) is estimated in gaussian notation as in (25)

$$I_{(r)} = I_0 e^{-\gamma r^2} \quad (25)$$

The firefly's brightness  $I$  shows its objective function's most recent position, as given by (26).

$$I_i = f(x_i) \quad (26)$$

Each firefly has an attractiveness value represented by  $\beta$ , and the less-bright firefly is drawn to the more-bright firefly. The formula for the variation of  $\beta$  with the distance,  $r$ , is stated in (27).

$$\beta_{(r)} = \beta_0 e^{-\gamma r^2} \quad (27)$$

where,  $\gamma$  = absorption coefficient of the media light,  $\beta_0$  = attractiveness value of firefly at  $r = 0$ .

### 5.4. Distance and Movement

The formula for the distance  $r_{ij}$  between the  $i^{th}$  and  $j^{th}$  fireflies, respectively located at  $x_i$  and  $x_j$ , is given by (28).

$$r_{ij} = \|x_i - x_j\| = \sqrt{\sum_{k=1}^d (x_{i,k} - x_{j,k})^2} \quad (28)$$

Where,  $x_i, k = k^{th}$  component of the spatial coordinate  $x_i$  of  $i^{th}$  firefly,  $d$  = distance.

If  $d$  equals 2, then (28) changes to (29).

$$r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (29)$$

The firefly  $i^{th}$  moves towards a more attractive firefly  $j^{th}$  as expressed by (29).

$$x_i^{t+1} = x_i^t + \beta_0 e^{-\gamma r_{ij}^2} (x_j^t - x_i^t) + \alpha (rand - 0.5) \quad (30)$$

Where, rand = random number within [0, 1],  $t$  = current iteration number,  $\alpha$  = value randomly selected often inside [0, 1],  $x_j$  = brighter firefly location,  $x_i$  = less bright firefly,  $\gamma$  = absorption coefficient and it lies in the range of 0.01 and 10.

The algorithm compares the new firefly attractiveness position value to the previous value. If the new site has a higher attraction rating than the old one, the firefly moves there; otherwise, it stays put. A predetermined fitness value determines the FA termination criterion. The brightest firefly will travel at random, according to (31).

$$x_i^{t+1} = x_i^t + \alpha \epsilon_i \quad (31)$$

The firefly will move randomly if there are no other fireflies around that are brighter. Up until the stopping condition is satisfied, the aforementioned procedures are repeated. The largest and best-predicted position and capacity are represented by the brightest firefly [25].

### 5.5. FA Implementation Algorithm for STATCOM Allocation

The following are the procedures in the Firefly algorithm for power flow incorporating a STATCOM controller.

- Enter the network data (independent parameters such as active power of all generators except the swing bus, generators' voltages, regulating transformer-tap setting, reactive power injection) while meeting different equality and inequality constraints.
- Initiate the firefly algorithm's parameters and constants, such as  $\alpha$ ,  $\beta_0$  and  $\gamma$ .
- Set the iteration count to 1 and generate 'n' fireflies at random.
- Execute the base case load flow.
- Use the mathematical formulation of the objective function in (14), to calculate the fitness function of each firefly for loss minimization.
- The fitness values are used to generate  $P_{best}$  values for all of the fireflies, with  $g_{best}$  being the best of the  $P_{best}$  values.
- Calculate each firefly's attraction distance by utilizing (29).
- For each firefly, new values are computed.
- Firefly's position is updated using (30).

- For each of the fireflies' new places, new fitness values are calculated. If a firefly's new fitness value is higher than its old  $P_{best}$  value, it is set to its current fitness value.  $G_{best}$  is calculated using the most recent  $P_{best}$  data.
- The iteration number is increased, and if it has not attained its maximum, the process proceeds to step 3 unless convergence is obtained.
- Sort the fireflies into categories based on the current global best. The optimal STATCOM capacities in 'n' candidates are determined by  $G_{best}$  firefly, with the position denoting the location and the results presented.

**6. Results and Discussion**

The load flow study findings, in addition to the applicability of the suggested PSO and FA for STATCOM controller optimum allocation to minimize losses and voltage violations of the IEEE 14-bus network, are shown. The control variables that were tuned include the voltage magnitude, tap parameters of the transformer, and STATCOM output. Table 1 shows these data for these control variables.

Accounting for the STATCOM power injection concept, MATLAB codes for a load flow study were written. These were employed for the load flow study in both cases—without and with the STATCOM controller. During the implementation and evaluation of both approaches on the IEEE 14-bus system, voltage profile augmentation and real as well as reactive power losses were employed as performance metrics.

Table 1: Restrictions on Control Parameters

S/N	Parameters	Limits
1	Voltage Magnitude	0.95 – 1.05 p.u
2	Tap Settings of the Transformer	0.90 – 1.10 p.u
3	Static Compensator MVar	0.00 – 100 MVar

**6.1. Voltage Profile**

The magnitudes of the network voltages are shown in Fig. 3. This implies that the bus voltage magnitudes were greatly enhanced following STATCOM regulator optimization using FA as opposed to when the PSO approach was used for the identical device setup. Bus voltage magnitudes across the entire test network are all within the permissible limits of 0.95 to 1.05 p.u., culminating in dependable network operation. The discrepancy was lessened by these two methods. FA did, however, provide the greatest voltage deviation minimization results in this circumstance.

Buses 2 and 3 offer a compelling justification for this performance. When optimizing with PSO, the bus 2 voltage was 1.048 p.u., and it was 1.046 p.u. after the controller was deployed with FA. The voltage magnitude at bus 3 increased from 0.96 to 0.98 p.u. and then to 0.99 p.u. as a result of optimization utilizing PSO and FA. Given that the anticipated voltage is 1.00 p.u., the best suitable method is one in which network influences attempt to return the voltage to that value.

**6.2. Minimization of Active Power Loss**

Utilizing optimization techniques for placement strategies, the FACTS controller decreased the active power loss. Fig. 4 depicts the active power loss data for both the PSO and FA-

placed STATCOM controllers, as well as the base case. The overall active loss for the base case was recorded at 6.251 MW. Applying PSO and FA to integrate the device reduced the loss to 5.819 and 5.518 MW. The overall loss was minimized by 0.733 MW following the device's incorporation using FA as opposed to 0.432 MW when PSO was employed.

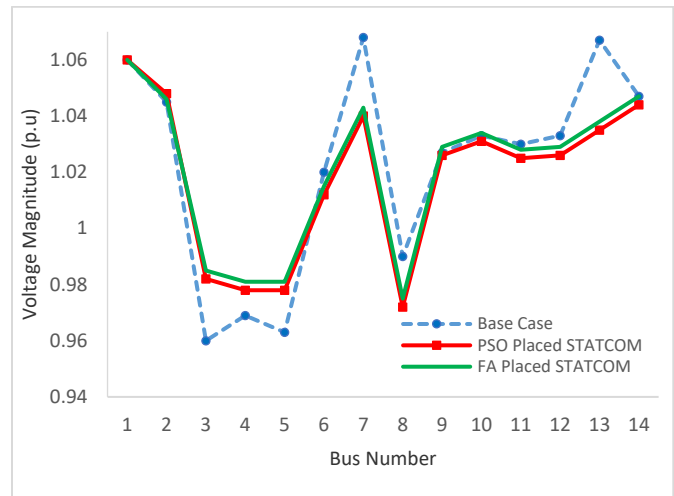


Figure 3: Voltage profile comparison

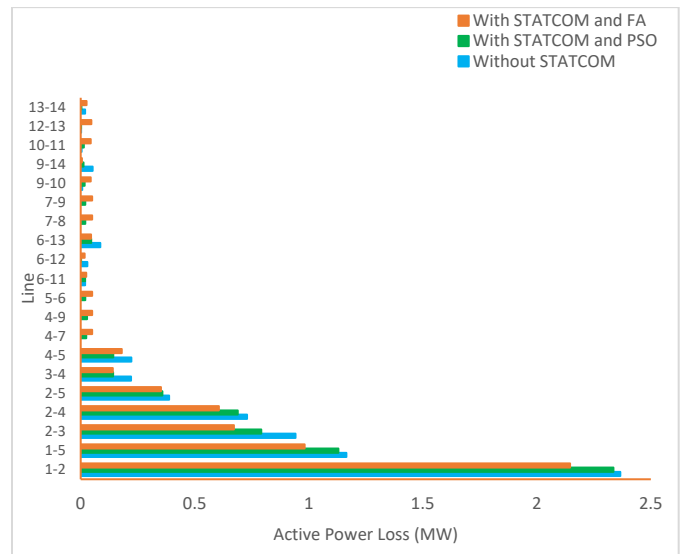


Figure 4: Active power loss minimization for all the cases

By rerouting the system load flow, the loss was reduced. PSO had a loss reduction of 6.9%, whereas FA had a loss reduction of 11.73%. This indicates that FA fared better than PSO in the active loss reduction of the system under study. A more thorough evaluation of the effectiveness of the two techniques in terms of loss reduction is also illustrated in Fig. 4. All of the lines displaying loss decreased following the controller installation utilizing the FA and PSO techniques. The degree of loss reduction does, however, differ between the two strategies. The green bars (loss with the PSO technique) have a substantially higher magnitude than the red bars (FA-placed STATCOM). As seen in the red illustration, these reductions with FA-placed STATCOM substantially outweigh those with

PSO placement. The overall loss minimization is shown in Fig. 5 to help understand how well PSO and FA may be used to optimize STATCOM controllers. It is impossible to exaggerate the advantages of FA over the PSO algorithm. Cost reductions were achieved as a consequence of FA's better minimization findings for active loss and voltage profile augmentation. The STATCOM controller's presence led to a redistribution of network power that improved network operation.

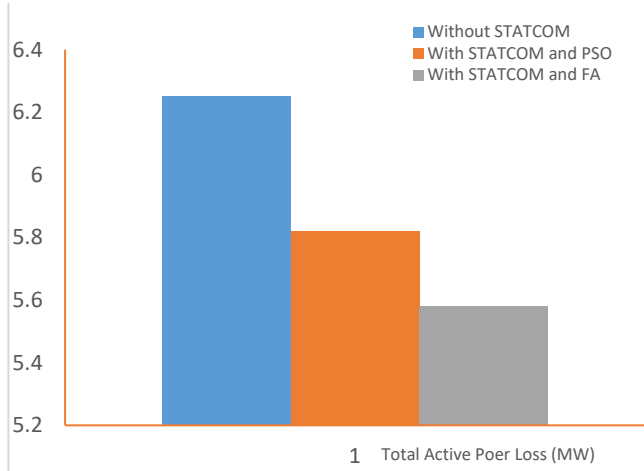


Figure 5: Overall active power losses

The controller offered a different flow path, allowing electricity to flow through less-loaded lines and reducing loss on the original lines as a consequence. The active power flows for the base case, which were 69.9246, 68.7359, 51.8444, 38.2318, 7.5796, 17.2293, 3.6629, and 5.4713 MW for transmission lines 1–5, 2-3, 4, 2, 5, 6, 12, 6, 13, and 14, were modified to 69.8589, 68.7203, 52.1509, 38.5039, 7.5424, 17.1875, and 3.5072 MW. On the other hand, the line flow was improved by 1.40, 1.04, 0.43, 0.08, 0.02, 0.09, 0.20, and 0.03 MW compared to the flow recorded with the PSO technique application.

The system's overall active power flow is therefore increased by utilizing these algorithms, from 621.5 to 623.4 MW with the PSO-placed STATCOM and to 626.64 MW with the FA-placed STATCOM.

### 6.3. Reduction of Reactive Power Loss

The findings of the network branch losses for the system under study, before and following STATCOM installation, are shown in Fig. 6. The overall power loss without the device was 14.256 MVar; nevertheless, when STATCOM's optimum configuration was attained with PSO, this decreased to 12.59 MVar. Following appropriate STATCOM integration using FA, this loss was further reduced to 12.16 MVar. The STATCOM device's integration with PSO and FA resulted in achievements of 1.62 and 2.10 MVar, or 11.37 and 14.73%, respectively, in overall reduction. When the two optimization techniques are compared for effectiveness, FA outperforms PSO in minimizing reactive power loss.

As illustrated in Fig. 6, all transmission lines—aside from lines 3–4—were loss-minimized utilizing FA-placed STATCOM. The disparities in reactive loss magnitude for appropriately located STATCOM with PSO and FA show that

FA has a loss reduction boost over PSO. The reduction of the system's overall reactive loss with and without correctly positioned STATCOM is illustrated in Fig. 7.

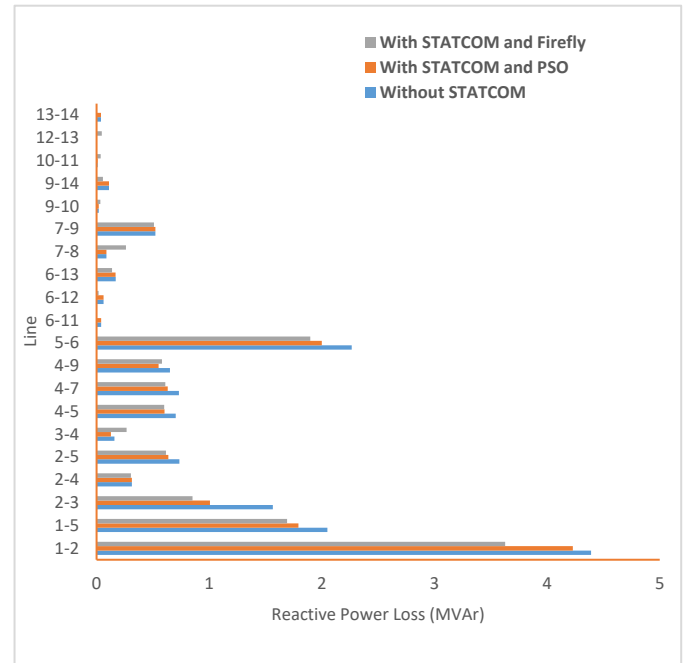


Figure 6: Reduction of reactive loss for all the cases

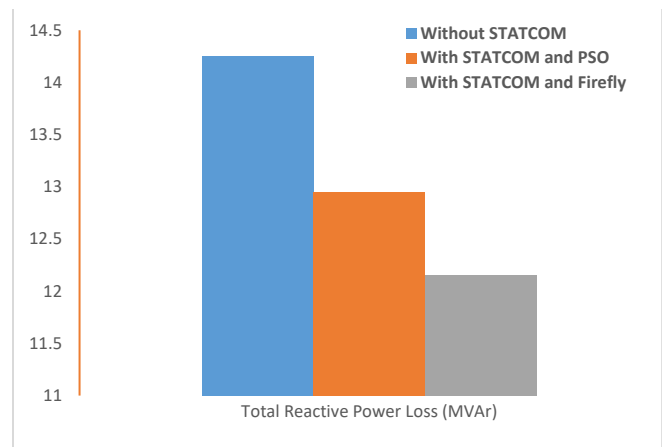


Figure 7: Overall reactive losses

This considerably decreased reactive loss on the network and outperformed the PSO technique. This reduction greatly aided in reducing the bus voltage magnitude deviation, which increased network stability and security. The test system's active and reactive loss information is shown in Table 2. Columns three and four, respectively, reflect the active and reactive power losses experienced by the network during the base case study. With PSO-placed STATCOM, the active power loss is recorded in column 5, while with FA-placed STATCOM, it is recorded in column 7. Columns six and eight of the table contain their related reactive power losses. As a consequence of the device using these two techniques, there is an overall line-by-line decrease for the active and reactive power.

Table 2: The IEEE 14-Bus Network Line Losses

Bus Number		Steady State (Base case)		STATCOM (PSO-placed)		STATCOM (Firefly-placed)	
From	To	(MW)	(MVar)	(MW)	(MVar)	(MW)	(MVar)
1	2	2.366	4.390	2.346	4.370	2.146	3.628
1	5	1.165	2.049	1.129	1.787	0.982	1.692
2	3	0.942	1.565	0.819	0.947	0.672	0.952
2	4	0.729	0.313	0.726	0.415	0.706	0.395
2	5	0.388	0.736	0.372	0.676	0.352	0.696
3	4	0.221	0.158	0.161	0.247	0.141	0.267
4	5	0.222	0.703	0.200	0.698	0.180	0.678
4	7	0.000	0.731	0.030	0.671	0.050	0.651
4	9	0.000	0.651	0.030	0.601	0.050	0.581
5	6	0.000	1.898	0.030	2.265	0.050	2.245
6	11	0.019	0.041	0.004	0.023	0.024	0.003
6	12	0.029	0.062	0.002	0.038	0.017	0.018
6	13	0.086	0.170	0.065	0.158	0.045	0.138
7	8	0.000	0.087	0.030	0.281	0.050	0.261
7	9	0.000	0.522	0.030	0.530	0.050	0.510
9	10	0.007	0.019	0.024	0.014	0.044	0.034
9	14	0.052	0.111	0.020	0.077	0.005	0.057
10	11	0.004	0.009	0.024	0.016	0.044	0.036
12	13	0.002	0.001	0.027	0.027	0.047	0.047
13	14	0.019	0.039	0.006	0.017	0.026	0.002
<b>Total</b>		6.251	14.256	5.819	12.954	5.681	12.891

Comparing with the PSO technique, FA's efficacy cannot be highlighted enough. With this performance, FA was able to minimize active and reactive power losses as well as voltage fluctuations more effectively, which reduced costs.

For better comprehension, Table 3 shows the entire system power flows and the corresponding total loss projections. Without a STATCOM device, Table 3 shows that the network

Table 3: The IEEE 14-Bus Network Line Losses Network Overall Power Flows and Losses

	Active and Reactive Power Flows			Active and Reactive Power Losses		
	Base Case	PSO-placed STATCOM	FA-placed STATCOM	Base Case	PSO-placed STATCOM	FA-placed STATCOM
Active (MW)	621.5	623.4	626.6	6.3	5.8	5.7
Reactive (MVar)	201.7	250.8	253.9	14.3	12.9	12.9
Apparent (MVA)	653.4	671.9	676.1	15.6	14.2	14.1

Table 4: STATCOM Parameters Settings and Location

Technique	Location	Voltage Value (p.u)	Angle (deg.)	STATCOM Size (MVar)
FA	9	1.029	0.926	9.54
PSO	11	1.025	3.769	8.96

## 7. Conclusion

This study looked into and proved the efficacy of the FA algorithm over the PSO method for placing STATCOM devices optimally. In this research study, the ideal STATCOM controller placement and parameter settings were made with the goals of reducing voltage magnitude variations and active and reactive power losses. The outcomes produced utilizing the IEEE 14-bus network show how appropriate these optimization strategies are. The capacity of the STATCOM controller to produce the best results for the specified objectives served as evidence of the applicability of PSO and FA for the best STATCOM controller position. According to the research, the STATCOM controller's performance with FA placement is superior to the PSO's. This means that FA performance in the optimum STATCOM

is capable of handling an apparent power of 653.38 MVA. But with the PSO and FA installed STATCOM controllers, the apparent power increased to 671.9 and 676.1 MVA, respectively. The system loss decreased from 15.57 to 14.20 and 13.81 MVA, respectively, as a result of this rise in total network power, as depicted in Table 3, when the device was strategically placed with PSO and FA, respectively.

Following the utilization of PSO and FA algorithms, the STATCOM allocation resulted in an improvement in overall flow of 2.8 and 3.5%, respectively. Deploying this device and employing the PSO and FA algorithms led to a reduction of the overall network loss of 8.78 and 11.26%, respectively. The stated FA performance in loss reduction and total network flow clearly demonstrates that FA is superior to PSO in the deployment of STATCOM device controllers. Table 3 indicates the differences in STATCOM device capacities.

Reactive power injection is represented by column 5, while STATCOM controller voltages and angles are shown in columns 3 and 4, respectively. Column 2 displays the device location that was selected. Table 4 shows the comparison of the total parameter settings and STATCOM controller location that led to the network performance for FA and PSO that was previously described. The table makes it evident that both algorithms' shunt reactances fall within the same range. As a consequence, the final device rating—which depends on the controller capacity and potential costs for the two techniques—is rather similar. Due to this, FA outperforms PSO in terms of cost.

controller configuration outperforms PSO in voltage profile augmentation and loss mitigation situations. Future research may be carried out to compare the effectiveness of the FA with other recently developed metaheuristic optimization algorithms that deliver superior performance at a reduced cost of STATCOM allocations on the power transmission and sub-transmission networks.

## Conflict of Interest

The authors declare no conflict of interest.

## References

- [1] E. I. Ogunwole, S. O. Ayanlade, D. E. Owolabi, A. Jimoh, A. B. Jimoh and F. M. Aremu, "Performance Comparative Evaluation of Metaheuristic Optimization Algorithms for Optimal Placement of Flexible Alternating Current Transmission System Device," in 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET), 2022, 1-8, doi: 10.1109/ICECET55527.2022.9872866.
- [2] B. Behera and K. C. Rout, "Comparative performance analysis of SVC, STATCOM UPFC during three-phase symmetrical fault," Proc. Int. Conf.

- Inven. Commun. Comput. Technol. ICICCT 2018, **2**, 1695–1700, 2018, doi: 10.1109/ICICCT.2018.8473226.
- [3] K. G. Damor, D. M. Patel, V. Agrawal, and H. G. Patel, "Comparison of different FACT devices," *Int. J. Sci. Eng.*, **1**(1), 372–375, 2014, doi: 10.1109/APCC.2016.7581484.
- [4] Jimoh, A., Ayanlade, S. O., Ariyo, F. K. and Jimoh, A. B., "Variations in phase conductor size and spacing on power losses on the Nigerian distribution network. *Bulletin of Electrical Engineering and Informatics*," **11**(3), 1222 - 1233, 2022, doi: <https://doi.org/10.11591/eei.v11i3.3753>.
- [5] E. I. Ogunwole, "Optimal placement of statcom controllers with metaheuristic algorithms for network power loss reduction and voltage profile deviation minimization," M.Tech Dissatation, Kwa Zulu Natal University, 2020.
- [6] B. O. Adewolu and A. K. Saha, "FACTS devices loss consideration in placement approach for available transfer capability enhancement," *Int. J. Eng. Res. Africa*, **49**, 104–129, 2020, doi: 10.4028/www.scientific.net/JERA.49.104.
- [7] S. O. Ayanlade and O. A. Komolafe, "Distribution system voltage profile improvement based on network structural characteristics," in OAU Faculty of Technology Conference (OAUTEKConf2019), 2019, 75–80.
- [8] A. K. Rawat et al., "Design of microcontroller based static VAR compensator," 2015 17th Eur. Conf. Power Electron. Appl. EPE-ECCE Eur. 2015, **4**(1), 1–6, 2017, doi: 10.1515/ijeeps-2017-0145.
- [9] A. Gupta and P. R. Sharma, "Optimal placement of FACTS devices for voltage stability using line indicators," in 2012 IEEE 5th Power India Conf. PICONF 2012, 4–6, 2012, doi: 10.1109/PowerI.2012.6479518.
- [10] S. T. Fadhil and A. M. Vural, "Comparison of dynamic performances of TCSC, STATCOM, SSSC on inter-area oscillations," in 2018 5th Int. Conf. Electr. Electron. Eng. ICEEE, 138–142, 2018, doi: 10.1109/ICEEE2.2018.8391317.
- [11] Y. Zhang, Y. Zhang, B. Wu, and J. Zhou, "Power injection model of STATCOM with control and operating limit for power flow and voltage stability analysis," *Electr. Power Syst. Res.*, **76**(12), 1003–1010, 2006, doi: 10.1016/j.epsr.2005.12.005.
- [12] D. Karaboga and B. Akay, "A comparative study of artificial bee colony algorithm," *Appl. Math. Comput.*, **214**(1), 108–132, 2009, doi: 10.1016/j.amc.2009.03.090.
- [13] M. Sankaramoorthy and M. Veluchamy, "A hybrid MACO and BFOA algorithm for power loss minimization and total cost reduction in distribution systems," *Turkish J. Electr. Eng. Comput. Sci.*, **25**(1), 337–351, 2017, doi: 10.3906/elk-1410-191.
- [14] A. Elansari, J. Burr, S. Finney, and M. Edrah, "Optimal location for shunt connected reactive power compensation," in *Proc. Univ. Power Eng. Conf.*, 1–6, 2014, doi: 10.1109/UPEC.2014.6934743.
- [15] M. O. Okelola, S. A. Salimon, O. A. Adegbola, E. I. Ogunwole, S. O. Ayanlade, and B. A. Aderemi, "Optimal siting and sizing of D-STATCOM in distribution system using new voltage stability index and bat algorithm," **2**(2), 2–6, 2021.
- [16] S. Majumdar, A. K. Chakraborty, and P. K. Chattopadhyay, "Active power loss minimization with FACTS devices using SA/PSO techniques," in 2009 Int. Conf. Power Syst. ICPS '09, 1–5, 2009, doi: 10.1109/ICPWS.2009.5442726.
- [17] A. A. Esmim and G. Lambert-Torres, "Loss power minimization using particle swarm optimization," in *IEEE Int. Conf. Neural Networks - Conf. Proc.*, 1988–1992, 2006, doi: 10.1109/ijcnn.2006.246945.
- [18] S. O. Ayanlade, E. I. Ogunwole, S. A. Salimon, and S. O. Ezekiel, "Effect of optimal placement of shunt facts devices on transmission network using firefly algorithm for voltage profile improvement and loss minimization," *Advances on Intelligent Informatics and Computing: Health Informatics, Intelligent Systems, Data Science and Smart Computing*, **127**, 385-396, 2022.
- [19] M. O. Okelola, S. O. Ayanlade, and E. I. Ogunwole, "Particle swarm optimisation for optimal allocation of STATCOM on transmission network," in *Journal of Physics: Conference Series*, 2021.
- [20] X. S. Yang and X. He, "Firefly algorithm: recent advances and applications," *Int. J. Swarm Intell.*, **1**(1), 36, 2013, doi: 10.1504/ijsi.2013.055801.
- [21] X. S. Yang, "Firefly algorithms for multimodal optimization," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5792 LNCS, 169–178, 2009, doi: 10.1007/978-3-642-04944-6\_14.
- [22] A. Ritthipakdee, A. Thammano, N. Premasathian, and D. Jitkongchuen, "Firefly mating algorithm for continuous optimization problems," *Comput. Intell. Neurosci.*, **2017**, 2017, doi: 10.1155/2017/8034573.
- [23] F. S. Moustafa, N. M. Badra, and A. Y. Abdelaziz, "Evaluation of the performance of different firefly algorithms to the economic load dispatch problem in electrical power systems," *Int. J. Eng. Sci. Technol.*, **9**(2), 1, 2017, doi: 10.4314/ijest.v9i2.1.
- [24] N. F. Johari, A. M. Zain, N. H. Mustaffa, and A. Udin, "Firefly algorithm for optimization problem," *Appl. Mech. Mater.*, **421**, 512–517, 2013, doi: 10.4028/www.scientific.net/AMM.421.512.
- [25] S. O. Ayanlade, E. I. Ogunwole, A. Jimoh, S. O. Ezekiel, D. E. Owolabi, and A. B. Jimoh, "STATCOM Allocation Using Firefly Algorithm for Loss Minimization and Voltage Profile Enhancement," in 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), 2022, 1-6, doi: 10.1109/ICECCME55909.2022.9988475.

## Northern Leaf Blight and Gray Leaf Spot Detection using Optimized YOLOv3

Brian Song<sup>1</sup>, Jeongkyu Lee<sup>2,\*</sup>

<sup>1</sup>Roslyn High School, Department, Institute, Roslyn Heights, 11577, USA

<sup>2</sup>Northeastern University, Khoury College of Computer Science, San Jose, CA, 95113, USA

### ARTICLE INFO

Article history:

Received: 01 September, 2022

Accepted: 21 January, 2023

Online: 24 February, 2023

Keywords:

Northern leaf blight

Gray leaf

Corn disease

YOLOv3

CBAM

### ABSTRACT

Corn is one of the most important agricultural products in the world. However, climate change greatly threatens corn yield, further increasing already prevalent diseases. Northern corn leaf blight (NLB) and Gray Leaf Spot are two major corn diseases with lesion symptoms that look very similar to each other, and can lead to devastating loss if not treated early. While early detection can mitigate the amount of fungicides used, manually inspecting maize leaves one by one is time consuming and may result in missing infected areas or misdiagnosis. To address these issues, a novel deep learning method is introduced based on the low latency YOLOv3 object detection algorithm, Dense blocks, and Convolutional Block Attention Modules, i.e., CBAM, which can provide valuable insight into the location of each disease symptom and help farmers differentiate the two diseases. Datasets for each disease were hand labeled, and when combined, the base YOLOv3, Dense, and Dense-attention had AP<sub>0.5</sub> NLB lesions/AP<sub>0.5</sub> Gray leaf spot lesions value pairs of 0.769/0.459, 0.763/0.448, and 0.785/0.483 respectively.

### 1. Introduction

Worldwide, 10 to 40% of crops die due to pests and diseases [1]. This issue will only get worse in the future, as temperature changes due to climate change will lead to more favorable conditions for pathogens [2]. In order to deal with this growing issue, it is necessary to come up with accurate diagnostic methods in order to quickly treat diseased plants. Focusing on maize, one of the most important crops in the world that accounts for two-thirds of the total volume of coarse grain trade globally in the past decade [3], in recent years, both Northern corn leaf blight, i.e., NLB, and Gray leaf spot disease has become more prevalent. In 2015, NLB was ranked first in most destructive corn disease in the northern United States and Ontario, Canada, up from seventh in 2012, with an estimated loss of 548 million bushels. For comparison, in 2015, the second-ranked most destructive corn disease, anthracnose stalk rot, had an estimated loss of 233 million bushels, less than half of NLB. In 2012, Gray leaf spot was ranked sixth in most destructive corn disease and has maintained its place as one of the top four most destructive from 2013-2015 [4]. NLB is caused by the fungus *Exserohilum turcium*, and the most distinguishing visual symptom of the disease is a cigar-shaped, tan lesion that can range from one to seven inches long, as shown in Figure 1. Gray leaf spot is a fungal disease caused by *Cercospora zea-maydis* with

rectangular lesions from two to three inches long, as shown in Figure 2., often leading to confusion by farmers due to its similarity to NLB lesions. Although fungicides can be used as a treatment for these two diseases, studies have shown that fungicides persist in aquatic systems and are toxic to organisms [5]. Early detection can mitigate the necessity of fungicides, traditionally through manual scouting [6]. However, this method is time-consuming and can result in inaccurate or missed diagnoses due to human error. As a result, several types of image-based machine learning solutions, such as convolutional neural networks, i.e., CNN, and object detection algorithms, have been proposed.



Figure 1: NLB infected maize images from Cornell CALS

CNNs are commonly used for image classification, which involves assigning an entire image to a single class label. However, in cases where the location of objects in an image is important,

\*Jeongkyu Lee, San Jose, CA, 95113, jeo.lee@northeastern.edu

image classification can be difficult to interpret and verify. Object detection algorithms, on the other hand, can identify the specific location and extent of objects in an image, and can even draw bounding boxes around them to highlight their location. These algorithms can be useful in situations such as identifying disease symptoms in natural environment images, where it is important to know the exact location of the symptom. The typical symptoms of NLB and Gray leaf spot, which are brown, oval lesions, can be difficult to distinguish with the naked eye, but computer vision algorithms may be able to identify and differentiate them.



Figure 2: Gray leaf spot infected maize images from Cornell CALS

Many different object detection algorithms have been created, such as YOLOv3, which is both accurate and has low inference speed. Low inference speed is essential to speed up traditional manual practices and increases the likelihood of detecting a lesion on live video. However, one of the tradeoffs for its high inference speed is reduced accuracy. As a result, an optimized algorithm based on YOLOv3 was proposed in this paper by applying two methods: Dense blocks and convolutional block attention modules, i.e., CBAM. These optimizations were chosen to increase accuracy while maintaining or reducing inference speed compared to the base YOLOv3. In addition to proposing an improved algorithm, one of the main challenges of object detection is the need for more high-quality datasets, especially in niche areas such as plant disease detection. It is much easier to create image classification datasets because the label for an entire image is a single class or word. For object detection datasets, each image may contain more than one class, numerous objects per class, and requires the tedious work of locating all objects in an image and drawing bounding box labels around them. As a result, if there are no object detection datasets for a specific class, datasets originally for CNNs may be used instead by converting them to the correct format. In summary, this paper offers the following contributions:

- Application of machine learning to detect NLB and gray leaf spot lesions
- An optimized YOLOv3 with improved detection ability without significantly increased inference speed
- A NLB and grey spot dataset suited for object detection

This paper is an extension of work originally presented in IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC) [7], and the paper is structured as follows: Section 2 is the related works. Section 3 explains the YOLOv3 model and its optimizations. Section 4 shows how the dataset was created and explains the evaluation metrics. Section 5 shows the performance of the algorithms, and section 6 is the conclusion.

## 2. Related Works

Both image classification algorithms, such as CNNs and object detection algorithms, have been successfully applied for plant disease diagnosis. In [8], the authors used a CNN for plant

disease diagnosis, training it on an extensive image dataset of close-up, individual diseased leaves using the Plant Village dataset. The dataset consists of images of twenty-six diseases of plants, such as those that affect corn and apple. They reported 99.35% accuracy on a held-out test set. However, the images in the dataset were taken in a lab environment, where variables such as the presence of multiple leaves in an image, orientation, brightness, and soil presence are not considered. Solving this issue, high-quality datasets have been created by experts. In [9], the authors took images of NLB-infected corn and annotated each lesion individually with line annotations, which is the same dataset used in this paper. Because the images covered a large area, the actual lesions took up a small portion of the overall image. As a result, using a CNN on scaled-down versions of the images resulted in 70% accuracy. It was only after dividing the image into grids and associating a diseased-or-not class to each grid using their line annotations that resulted in 97.8% accuracy. Although successful, the process of dividing the image and then running inferences finally resulting in a high accuracy indicates a potential limitation of CNNs – they are not suited when the characteristics that define a class are small compared to the entire image.

On the other hand, object detection algorithms are suited to looking for specific areas in an image. The single shot detector, i.e., SSD [10] object detection algorithm, was used to detect apple diseases such as Brown Spot and Grey spot [11]. Instead of labeling an entire infected leaf as belonging to a class, the author only labeled the specific symptoms, such as spots, allowing the SSD algorithm to learn that the presence of a particular cluster of pixels determines the final output. The YOLOv3 [12] object detection algorithm was used to locate characteristics of various tomato diseases, such as early blight and mosaic disease [13]. The authors collected their own tomato dataset and annotated them by grouping clusters of diseased symptoms. In [14], the authors used a variant of an SSD and the Faster R-CNN [15] algorithm for grape plant disease object detection. They used existing datasets such as the Plant Village dataset, which mentioned previously is used for CNN training, drew boxes around the grape disease, converting it for object detection use. However, instead of drawing boxes around the grape disease symptoms, they label the entire leaf containing the disease, meaning it could be more specific. In [16], the authors also annotated the Plant Village dataset with bounding boxes for object detection. Instead of limiting to only certain disease classes, the entire dataset was used, resulting in about 54,000 annotated images. In [17], the authors created their object detection dataset called PlantDoc, which includes corn diseases, including both NLB and gray leaf spot. However, the dataset is small, less than 200 images per class, and the annotations are not very specific – clusters of lesions are grouped. This leads to the additional problem of determining whether two lesions belong to the same or different clusters, especially in images where lesions make up most of the corn leaf.

## 3. Methods

In this section, deep learning methods such as the usage of dense blocks, CBAM, and the proposed optimized algorithm to detect corn disease will be introduced.

### 3.1. Base Algorithm: YOLOv3

YOLOv3 is an object detection algorithm that boasts high accuracy without sacrificing speed. It is an improvement of YOLO



in 2016 [18] and YOLOv2 in 2017 [19]. Although newer variants exist, such as YOLOv4 [20], for this study, YOLOv3 was chosen as the algorithm to focus on because there is many open source code for the algorithm, meaning it was easier to find a repo with an implementation that could easily be modified. Also, YOLOv3, as shown in Figure 3, uses the darknet-53 backbone to extract features of images. The backbone network utilizes residual blocks containing skip connections, which was introduced in ResNet [21]. These skip connections skip some layers in the backbone, helping to alleviate the vanishing gradient problem and making it easier to tune the earlier layers of a network. In the figure, the residual  $N$  blocks consist of a  $3 \times 3$  convolutional layer and  $N$  residual units. The detection stages contain additional convolutional layers for further feature extraction and detect potential objects on three different scales. These scales are used to detect large, medium, and small-sized objects. This improves the performance of varying image sizes. According to the YOLOv3 paper, performance on the Microsoft Common Objects in Context, i.e., MS COCO, dataset, a benchmark used for evaluating object detection algorithms in which the accuracy metric mean average precision, i.e.,  $mAP$ , is commonly used, showed that the three fastest were YOLOv3-320, SSD321, and DSSD321 [22] with 51.5  $mAP$ -50/22 ms, 45.4  $mAP$ -50/61 ms, and 46.1  $mAP$ -50/85 ms, respectively, indicating  $mAP$  and inference time. Compared to the other algorithms, YOLOv3 is significantly faster while achieving better  $mAP$ . As a result, YOLOv3 was selected as the base algorithm of our research because efficiency is critical for searching through large cornfields.

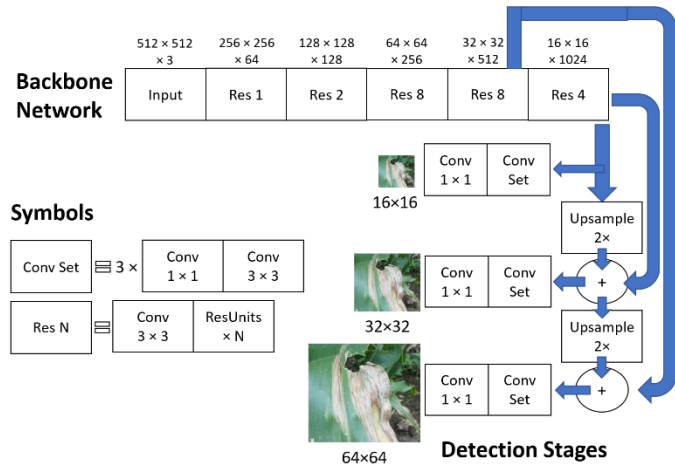


Figure 3: YOLOv3 diagram

### 3.2. Dense Block

DenseNet [23] is a convolutional neural network architecture that uses dense blocks, in which the output of each convolution layer is connected to the inputs of all subsequent layers. This allows later layers to use information learned in earlier layers and reduces the number of parameters, improving computational efficiency and mitigating the vanishing gradient problem. Transition layers, which consist of a  $1 \times 1$  convolution and average pooling layer, are placed between groups of dense blocks to reduce the number of parameters and dimensionality. The growth rate, denoted by  $k$ , determines the number of new feature maps added for each layer in a dense block.

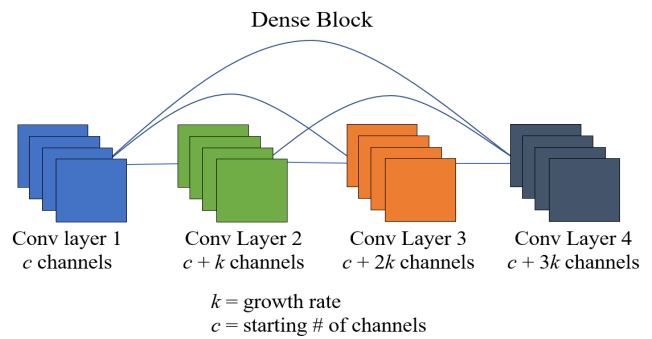


Figure 4: Diagram of 4 layer dense block

### 3.3. CBAM

CBAM [24] is a method that uses the attention mechanism to replicate how humans pay attention to their environment. It uses both channel and spatial attention to focus on particular objects in a scene and enhance the feature maps of the convolutional layers. Channel attention selects the most important channels and weighs them to improve them, while spatial attention applies max pooling and average pooling to help the model know where to focus in the image. Using this method leads to improved accuracy with only a limited amount of extra computation.

### 3.4. Proposed Algorithm

The proposed algorithm for an input image of  $512 \times 512$  pixels is shown in Figure 5. The algorithm includes changes to the base YOLOv3 model, indicated by the yellow and green shaded portions. After the third residual block, a four-layer dense block is inserted, followed by a transition layer which reduces the number of filters and dimensions by half. The fourth and fifth residual blocks of the original backbone network are replaced with two six-layer dense blocks, with another transition layer in between. The growth rate,  $k$ , for all dense blocks is set to 128 to create a wide and shallow network for increased speed efficiency. CBAM blocks, which implement the attention mechanism, are placed at the beginning of the detection stages to refine the final feature maps and improve accuracy. The entire proposed model is called the Dense-attention algorithm, while the Dense algorithm is the same but without the CBAM blocks.

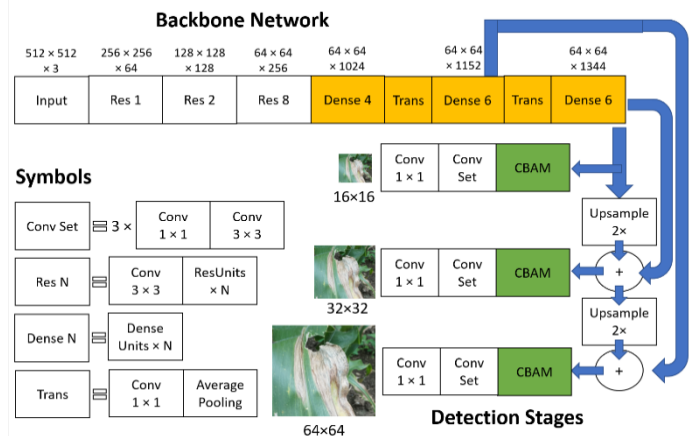


Figure 5: Proposed algorithm diagram

## 4. Experimental Setup

### 4.1. Datasets

To evaluate the proposed method in this paper, two datasets, one for NLB and the other for gray leaf spot, are combined. For the first dataset, the *Field images of maize annotated with disease symptoms dataset* [25] was used. The dataset consists of natural environment 4000×6000 pixel images of maize leaves taken by hand during the 2015 growing period in Aurora, New York. The main axis of the NLB lesions in each image is annotated with line annotations. 376 images are extracted. Because of the high image resolutions, each image is split into a 2×3 grid for pixel preservation. For the original line annotations, multiple lines would be used to signify one contingent lesion when only one label should be used. As a result, annotations are reconverted into bounding box format by hand under the supervision of the original line annotations. Figure 6 shows the annotation process. Images without lesions are discarded, resulting in 999 final images. Data augmentation using random rotation and zoom was then used to 4x the dataset size.

Unlike for NLB, high quality datasets for gray leaf spot disease taken from the natural environment and annotated by experts are rare. Although an annotated gray leaf spot dataset from Plant Doc exists, as mentioned in the related works section, few images are provided. Additionally, clusters of spots are grouped as one label, as opposed to the individual disease symptom labeling that the format of the NLB dataset is in. As a result, the Plant Village dataset was chosen to be annotated, starting from scratch. Because the dataset was initially intended for CNN training, annotations are not provided. As a result, like the NLB dataset, new annotations around each spot were created. However, in the gray leaf spot dataset, no expert guidance was provided, meaning best guesses for what constituted a spot were made. Figure 7 shows examples of grey spot annotations. Although fewer images of gray leaf spot were used compared to NLB, more annotations per image for gray were created, balancing the total number of annotations per class. The Plant Village dataset, in addition to the gray leaf spot, contains healthy corn images. 200 healthy images from each dataset were added. Several dataset groupings were used in this study.

Table 1: Dataset description

Dataset	Training images	Validation images	Testing images	Total label count
NLB	3,145	135	135	5,255
Gray	608	24	30	5,677
NLB+Gray	3,753	159	165	10,932
Healthy	4,033	219	225	10,932

Dataset NLB: NLB infected images only, Dataset Gray: Gray leaf spot infected images only, Dataset NLB+Gray: only NLB and gray leaf spot images, and finally, Dataset Healthy: a combination of dataset NLB+Gray with healthy images. Healthy images from the original NLB and Plant Village datasets were also included so that the algorithms could better learn what is not considered diseased through more examples. More info regarding dataset size and label counts is shown in Table 1. Data is available at a project github site<sup>a</sup>.



Figure 6: NLB annotation process



Figure 7: Grey leaf annotation examples

### 4.2. Evaluation Metrics

Evaluating accuracy for object detection algorithms requires comparing both the label and location of the predicted bounding box to those of the ground truth. *IoU*, i.e., intersection over union, as given in equation (1), is the value of the intersection area of the bounding box,  $Area_{pred}$ , and the area of the ground truth box,  $Area_{truth}$ , over the union area of the two aforementioned boxes.

$$IoU = \frac{Area_{pred} \cap Area_{truth}}{Area_{pred} \cup Area_{truth}} \quad (1)$$

If *IoU* is above a certain threshold value, that prediction is counted as a true positive. If not, it is considered a false positive. As shown in equation (2), precision, or PR, is the number of true positives divided by the sum of the number of true positives and false positives. Recall, or RE, shown in equation (3), is the number of true positives divided by the true positives and false negatives.

$$PR = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (2)$$

$$RE = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (3)$$

Average precision, i.e., AP, is calculated by finding the area under the precision-recall curve, shown in formula (4).  $P(r)$  is the precision value at recall value  $r$ . AP subscript  $k$  indicates the average precision when *IoU* is at threshold  $k$ .  $AP_{0.5}$  indicates average precision at the 0.5 *IoU* threshold. If three lesions had *IoU* of 0.2, 0.6, and 0.3, only the lesion with the *IoU* 0.6 would be counted as a true positive.

$$AP = \int_0^1 p(r) dr \quad (4)$$

Inference speed, parameter count, and MFLOPs were also measured. MFLOPs is a unit for how many million floating point operations per second the computer can operate. All results were

<sup>a</sup><https://github.com/beans1321/NLB-Grey-dataset/tree/main>

tested on the same GPU and Google colab environment settings for fair comparisons.

### 5. Results

Model training and evaluation were done on colab with an Nvidia P100 GPU, Intel Xenon 2.2 GHz CPU, and 13 GB of RAM. Code implementations were done using TensorFlow 1.15.0 and Keras 2.1.6 with Python 3.7. The three different algorithms, (i) *Base*, (ii) *Dense*, and (iii) *Dense-attention* were trained and evaluated on images sized 512×512 pixels.

Table 2 shows the  $AP_{0.5}$  of the algorithms on the datasets. On the NLB dataset, *Base*, *Dense*, and *Dense-attention* had  $AP_{0.5}$  of 0.774, 0.806, and 0.821, respectively. In the Gray dataset, *Base*, *Dense*, and *Dense-attention* had  $AP_{0.5}$  of 0.484, 0.471, and 0.496, respectively, showing that locating the exact boundaries of the grey spot is much more complicated than finding NLB lesions. One of the possible reasons for this is that NLB may appear more distinctive in the images since the lesions in the images tend to appear brighter than lesions in the Gray dataset. In the NLB+Gray dataset, individual  $AP_{0.5}$  was recorded for each class. The  $AP_{0.5}$  for each class decreased slightly compared to detecting them in dataset NLB and dataset Gray, in which only one class was present in each, which is expected as multiclass detection is more difficult than single class detection. The slight decrease in  $AP_{0.5}$  also indicates that the model has learned to differentiate NLB and Gray leaf spot. In the Healthy dataset, the usage of images of healthy images decreased performance. All results for healthy were worse than the results for the NLB+Gray dataset. *Base* outperformed *Dense* and *Dense-Attention* in terms of finding NLB lesions, with  $AP_{0.5}$  of 0.714, 0.675, and 0.702, respectively. However, in terms of finding gray lesions, *Dense-attention*'s  $AP_{0.5}$  of 0.473 was still higher than *Base*'s  $AP_{0.5}$  of 0.425. One possible reason that adding healthy images did not help performance is that the healthy images may have simply included more objects that looked like lesions, such as a dry or dead leaf, which are common, making training harder for the models.

Table 2: Performance of algorithms measured in  $AP_{0.5}$

Dataset	Base	Dense	Dense-attention
NLB	0.774	0.806	0.821
Gray	0.484	0.471	0.496
NLB lesions in NLB+Gray	0.769	0.763	0.785
Gray lesions in NLB+Gray	0.459	0.448	0.483
NLB lesions in Healthy	0.714	0.675	0.702
Gray lesions in Healthy	0.425	0.428	0.473

Examples of detections are shown in Figure 8 and 9. Figure 8 shows detections of NLB by the *Dense-attention* trained from the NLB+Gray dataset. Figure 9 shows detections of Grey spot, also by *Dense-attention* trained from NLB+Gray. In Figure 9, it can be shown that the model has difficulty determining if two lesions

close to each other are one or two lesions, as shown in the top row. In the bottom row, it can be shown that the model also has difficulty finding very small lesions.



Figure 8: NLB detection of *Dense-attention* trained on NLB+Gray dataset

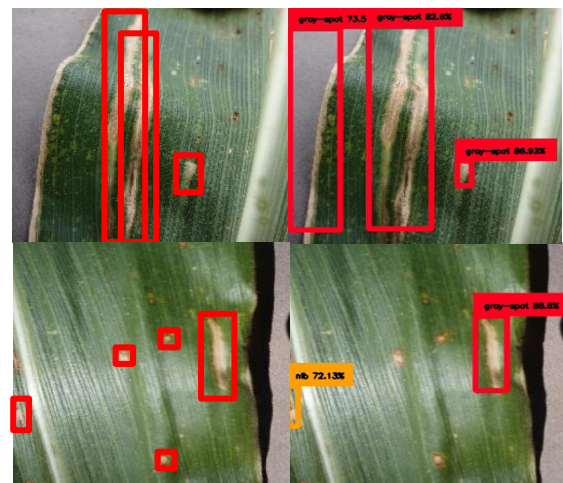


Figure 9: Gray detections of *Dense-attention* (right column) trained on NLB+Gray dataset side by side with ground truth (left column)

Table 3: General algorithm performance

Algorithm	Inference Speed (ms)	Parameters (millions)	MFLOPs
Base	36.9	61.5	123.0
Dense	37.4	40.8	81.6
Dense-attention	39.1	40.9	81.7

Table 3 shows the general performance of the algorithms, which shows constant results for the algorithms, no matter what dataset is used. *Base* has the fastest inference speed of 36.9 ms, followed by *Dense* with a speed of 37.4 ms, and finally *Dense-attention*, with a speed of 39.1 ms. There is not much of a major difference in inference speeds among all three algorithms. The usage of dense blocks to replace several layers of the original YOLOv3 backbone has greatly reduced the number of parameters. *Base* has 61.5 million parameters, *Dense* has 40.8 million, and *Dense-attention* has 40.9 million. The usage of CBAM has only increased the number of parameters slightly. The ratios of the parameter count between the algorithms are similar to the ratio of MFLOPs. *Base* has 123.0 MFLOPs, *Dense* 81.6, and *Dense-attention* 81.7. While *Base* is the fastest, *Dense* and *Dense-attention* are similar in speed while having drastically

fewer parameters and MFLOPs. In terms of accuracy, *Dense* seems to have similar results with *Base*, although it was more accurate in the NLB dataset. *Dense-attention*, based on the results shown in Table 2, mostly outperforms both *Base* and *Dense-attention*.

## 6. Conclusion

The new proposed model, called dense-attention, was built off of YOLOv3 and optimized for both accuracy and speed. New datasets for NLB and gray leaf spot were created and reannotated to be more suitable for object detection tasks. The results showed that dense-attention outperformed the base model in terms of accuracy, parameter count, and computational efficiency, although it was slightly slower. When both NLB and gray leaf spot were combined in the dataset, performance for each class decreased slightly compared to training on just one of the diseases. This suggests that the model was able to distinguish between the two visually similar diseases. In future work, it may be helpful to annotate the gray leaf spot dataset with experts and to include wider views of diseased leaves in the dataset.

## References

- [1] S. Savary, L. Willocquet, S. J. Pethybridge, P. Esker, N. McRoberts, A. Nelson, "The global burden of pathogens and pests on major food crops," *Nature Ecology & Evolution*, **3**(3), 430–439, 2019, doi:10.1038/s41559-018-0793-y.
- [2] T. M. Chaloner, S. J. Gurr, D. P. Bebber, "Plant pathogen infection risk tracks global crop yields under climate change," *Nature Climate Change*, **11**(8), 710–715, 2021, doi:10.1038/s41558-021-01104-8.
- [3] M. McConnell, "Feedgrains sector at a glance," USDA Economic Research Service U.S. Department of Agriculture, 27 January 2023, <https://www.ers.usda.gov/topics/crops/corn-and-other-feed-grains/feed-grains-sector-at-a-glance/>.
- [4] D. S. Mueller et al., "Corn yield loss estimates due to diseases in the United States and Ontario, Canada from 2012 to 2015," *Plant Health Progress*, **17**(3), 211–222, 2016, doi:10.1094/PHP-RS-16-0030.
- [5] J. P. Zubrod et al., "Fungicides: An Overlooked Pesticide Class?," *Environmental Science & Technology*, **53**(7), 3347–3365, 2019, doi:10.1021/acs.est.8b04392.
- [6] K. Wise, "Northern corn leaf blight - purdue extension," Purdue University, 2011, <https://www.extension.purdue.edu/extmedia/BP/BP-84-W.pdf>.
- [7] B. Song, J. Lee, "Detection of Northern Corn Leaf Blight Disease in Real Environment Using Optimized YOLOv3," *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, 2022, 0475-0480, doi:10.1109/CCWC54503.2022.9720782.
- [8] S. P. Mohanty, D. P. Hughes, M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, **7**, 2016, doi:10.3389/fpls.2016.01419.
- [9] C. DeChant, T. Wiesner-Hanks, S. Chen, E. L. Stewart, J. Yosinski, M. A. Gore, R. J. Nelson, H. Lipson, "Automated identification of northern leaf blight-infected maize plants from field imagery using Deep Learning," *Phytopathology*, **107**(11), 1426–1432, 2017, doi:10.1094/PHYTO-11-16-0417-R.
- [10] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg, "SSD: Single shot multibox detector," *Computer Vision–ECCV 2016*, 21–37, 2016, doi:10.1007/978-3-319-46448-0\_2.
- [11] P. Jiang, Y. Chen, B. Liu, D. He, C. Liang, "Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks," *IEEE Access*, vol 7, 59069–59080, 2019, doi:10.1109/ACCESS.2019.2914929.
- [12] J. Redmon, A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv*, 2018, doi:10.48550/ARXIV.1804.02767.
- [13] J. Liu, X. Wang, "Tomato diseases and pests detection based on improved Yolo v3 convolutional neural network," *Frontiers in Plant Science*, **11**, 2020, doi:10.3389/fpls.2020.00898.
- [14] S. Ghoury, C. Sungur, A. Durdu, "Real-Time Diseases Detection of Grape and Grape Leaves using Faster R-CNN and SSD MobileNet Architectures," *International Conference on Advanced Technologies, Computer Engineering and Science (ICATCES 2019)*, 39-44, 2019 Alanya, Turkey.
- [15] S. Ren, K. He, R. Girshick, J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**(6), 1137–1149, 2017, doi:10.1109/TPAMI.2016.2577031.
- [16] M. H. Saleem, S. Khanchi, J. Potgieter, K. M. Arif, "Image-Based Plant Disease Identification by Deep Learning Meta-Architectures," *Plants*, **9**(11), 1451, 2020, doi: 10.3390/plants9111451.
- [17] D. Singh, N. Jain, P. Jain, P. Kayal, S. Kumawat, N. Batra, "Plantdoc," *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, 2020, doi:10.1145/3371158.3371196.
- [18] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, doi:10.1109/CVPR.2016.91.
- [19] J. Redmon, A. Farhadi, "YOLO9000: Better, Faster, Stronger," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6517-6525, 2017, doi:10.1109/CVPR.2017.690.
- [20] A. Bochkovskiy, C.-Y. Wang, H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv*, 2020, doi:10.48550/ARXIV.2004.10934.
- [21] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778, 2016, doi: 10.1109/CVPR.2016.90.
- [22] C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi, A. C. Berg, "DSSD: Deconvolutional Single Shot Detector," *arXiv*, 2017, doi: 10.48550/ARXIV.1701.06659
- [23] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, "Densely Connected Convolutional Networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261–2269, 2017, doi:10.1109/CVPR.2017.243.
- [24] S. Woo, J. Park, J.-Y. Lee, I. S. Kweon, "CBAM: Convolutional Block Attention Module," *Proceedings of the European Conference on Computer Vision (ECCV)*, 11211, 3-19, 2018, doi: 10.1007/978-3-030-01234-2\_1.
- [25] T. Wiesner-Hanks, E. L. Stewart, N. Kaczmar, C. DeChant, H. Wu, R. J. Nelson, H. Lipson, M. A. Gore, "Image set for Deep Learning: Field images of maize annotated with disease symptoms," *BMC Research Notes*, **11**(1), 2018, doi:10.1186/s13104-018-3548-6.

## Active Simulation of Grounded Parallel-Type Immittance Functions Employing VDBAs and All Grounded Passive Components

Pratya Mongkolwai<sup>1</sup>, Pitchayanin Moonmuang<sup>2</sup>, Worapong Tangsrirat<sup>2,\*</sup>, Taweepol Suesut<sup>2</sup>

<sup>1</sup>Department of Instrumentation Engineering, Faculty of Engineering, Rajamangala University of Technology Rattanakosin, Nakhon Pathom 73170, Thailand

<sup>2</sup>Department of Instrumentation and Control Engineering, School of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

### ARTICLE INFO

Article history:

Received: 08 October, 2022

Accepted: 15 January, 2023

Online: 24 February, 2023

Keywords:

Voltage Differencing Buffered

Amplifier (VDBA)

Immittance function

Impedance simulator

### ABSTRACT

This communication proposes a grounded immittance function simulator that, depending on the proper choice of the passive components, can simulate parallel-type impedances of the R-L, R-C, and L-C forms. Only two grounded passive components and two voltage differencing buffered amplifiers (VDBAs) are used to implement the suggested circuit. All three simulated equivalent elements, namely  $R_{eq}$ ,  $L_{eq}$ , and  $C_{eq}$ , can be electronically adjusted through the VDBA's transconductance gain. The impact of the non-ideality of the VDBA device on the developed simulator is examined in detail. The voltage-mode bandpass filter has been implemented using the suggested active LC parallel impedance simulator to show that it performs as predicted. To prove the theory, the proposed circuit is simulated using the PSPICE tool. The findings of the experimental measures are also presented to demonstrate the circuit's feasibility.

## 1. Introduction

Electronic devices have assimilated into our daily lives in the world today. The development of novel technologies will be influenced by the published findings. In several analog signal processing solutions, the different active devices, such as current conveyor (CC), operational transconductance amplifier (OTA), current feedback operational amplifier (CFOA), and current differencing buffered amplifier (CDBA), have gained widespread attention. Similarly, since 2008, the voltage differencing buffered amplifier (VDBA) has been recognized as one of the most versatile and practical devices [1]-[2].

The VDBA element has a tunable transconductor as the input section and a voltage buffer as the output section. Because of this feature, this active element can be used in a variety of voltage-mode, current-mode, and mixed-mode analog circuits and applications [2-6]. Passive elements, such as resistors, capacitors, and inductors, were used in a variety of applications, including analog active filter circuits, sinusoidal oscillator design, and impedance cancellation circuit. However, when applied in the implementation of an integrated circuit (IC), the behavior of

passive elements was constrained by its enormous size and suffered from electronic tuning properties. As a consequence, an IC that mimicked the behavior of a passive element was implemented using an active element [7-9]. The parallel-type R-L simulators that were suggested in the literature [10-12] needed at least three active components. Similar to that, three or more passive components are required to realize the circuits in [11-12]. The circuits in [13] also need a high-voltage operation.

Therefore, the contribution of this work is to propose a grounded parallel R-L, R-C, and L-C impedance simulator, which depends on the appropriate selection of the passive element being used. The suggested simulator circuit uses only two VDBAs, two grounded passive components, and allows electronically control of the equivalent simulated elements via the transconductance gains of the VDBAs. In this study, the VDBA non-ideality effect on the actual immittance simulator is examined. With 0.18- $\mu\text{m}$  CMOS technology, the proposed R-L, R-C, and L-C impedance simulator circuit in frequency domain was simulated using PSPICE program. Time-domain analysis and temperature-dependent simulation are also carried out in the parallel R-C simulator. The theoretical analysis is validated by experimental laboratory measurements using commercially available IC LT1228. Additionally, the active L-C simulator has also been used to apply a second-order voltage-

\*Corresponding Author: Worapong Tangsrirat, Email: worapong.ta@kmitl.ac.th

mode bandpass filter in order to validate the viability. All results, both from simulations and experiments, are discovered to be in accordance with the theoretical predictions.

## 2. Fundamental of VDBA

Figure 1 depicts the electrical symbol of the VDBA element. This functional block has two input terminals (p and n) that meet high input impedance criteria and two output terminals (z and w), which have high and low impedances, respectively. Under ideal operating condition, the effective transconductance gain ( $g_m$ ) of the VDBA converts the differential voltage between  $v_p$  and  $v_n$  ( $v_p - v_n$ ) into an output current ( $i_z$ ) at terminal z. The voltage drop ( $v_z$ ) at the z terminal is transferred to the output voltage ( $v_w$ ) at the w terminal. From its ideal operating condition, the following matrix equation can be used to characterize the terminal relationship of the VDBA [1-2]:

$$\begin{bmatrix} i_p \\ i_n \\ i_z \\ v_w \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ g_m & -g_m & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} v_p \\ v_n \\ v_z \\ i_w \end{bmatrix} \quad (1)$$

In general, the  $g_m$  value in (1) can be changed by electronic means via the external bias voltage or current.

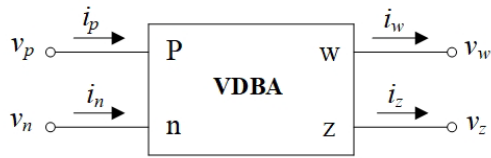


Figure 1: Schematic symbol of the VDBA.

In non-ideal assumption, the characteristic of VDBA can be modified as [3]:

$$\begin{bmatrix} i_p \\ i_n \\ i_z \\ v_w \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \alpha g_m & -\alpha g_m & 0 & 0 \\ 0 & 0 & \beta & 0 \end{bmatrix} \begin{bmatrix} v_p \\ v_n \\ v_z \\ i_w \end{bmatrix} \quad (2)$$

In above expression,  $\alpha = (1 - \varepsilon_{gm})$  and  $\beta = (1 - \varepsilon_v)$ , where  $|\varepsilon_{gm}| \ll 1$  and  $|\varepsilon_v| \ll 1$  stand for transconductance inaccuracy coefficient and the voltage tracking error, respectively.

A CMOS model of VDBA consisting of the differential amplifiers with active load ( $M_1$ - $M_4$  and  $M_7$ - $M_{10}$ ), and the source follower ( $M_{11}$ ) is shown in Figure 2. For the CMOS VDBA in Figure 2, the relationship between  $g_m$  and the bias current  $I_B$  can be characterized as follows [4]:

$$g_m = \sqrt{\mu C_{ox} \left( \frac{W}{L} \right) I_B} \quad (3)$$

Here,  $\mu$  is the effective carrier mobility,  $C_{ox}$  is the gate-oxide capacitance per unit area, and  $W$  and  $L$  are the effective channel width and length of  $M_1$  and  $M_2$  transistors, respectively.

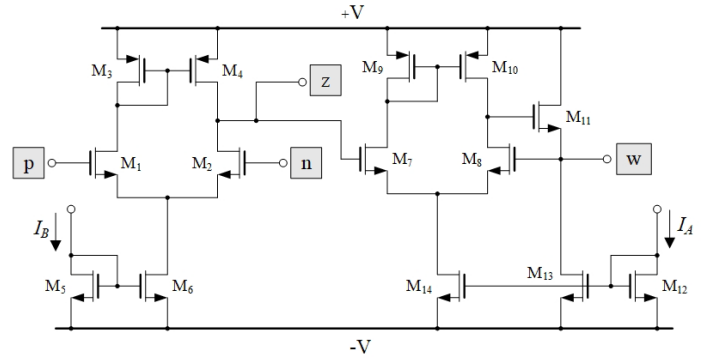


Figure 2: CMOS model of the VDBA used in this work.

## 3. Proposed Parallel-Type Immittance Function Simulator

According to Figure 3, the suggested grounded parallel-type immittance simulator is made up of two VDBAs and two grounded passive components. Based on ideal condition consumption, the input admittance ( $Y_{in}$ ) of the circuit is derived as:

$$Y_{in} = \frac{i_{in}}{v_{in}} = g_{m1} g_{m2} Z_A + \frac{1}{Z_B} \quad (4)$$

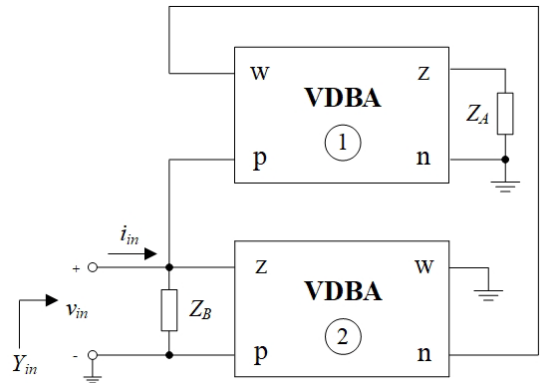


Figure 3: Proposed grounded parallel-type immittance function simulator.

The proposed parallel R-L, R-C, and L-C immittance simulator was made achievable by selecting the appropriate passive components, which describes the realized circuit. Its simulated impedances are summarized in Table 1, which illustrates that all synthetic simulator values can electronically be changed by the transconductance  $g_{mi}$  of the  $i$ -th VDBA ( $i = 1, 2$ ). Since all of the passive components are grounded, the configuration is attractive from further integration point of view. Another attractive feature of the design is that it does not need any special component equality for its realization.

Under the non-ideal operation given in (2), the results of reevaluating the proposed circuit in Figure 3 can be summarized in Table 2.

Table 1: Equivalent Circuit and Corresponding Equivalent Values for Figure 3 in Ideal Case

$Z_A$	$Z_B$	Equivalent circuit	Equivalent values
$1/sC_A$	$R_B$	parallel R-L	$R_{eq} = R_B$ ,

			$L_{eq} = \frac{C_A}{g_{m1}g_{m2}}$
$R_A$	$1/sC_B$	parallel R-C	$R_{eq} = \frac{1}{R_A g_{m1} g_{m2}}$ , $C_{eq} = C_B$
$1/sC_A$	$1/sC_B$	parallel L-C	$L_{eq} = \frac{C_B}{g_{m1} g_{m2}}$ , $C_{eq} = C_A$

Table 2: Equivalent Element Values for Figure 3 in Non-Ideal Case

$Z_A$	$Z_B$	Equivalent circuit	Equivalent values
$1/sC_A$	$R_B$	parallel R-L	$R_{eq} = R_B$ , $L_{eq} = \frac{C_A}{\alpha_1 \alpha_2 \beta_1 g_{m1} g_{m2}}$
$R_A$	$1/sC_B$	parallel R-C	$R_{eq} = \frac{1}{\alpha_1 \alpha_2 \beta_1 R_A g_{m1} g_{m2}}$ , $C_{eq} = C_B$
$1/sC_A$	$1/sC_B$	parallel L-C	$L_{eq} = \frac{C_B}{\alpha_1 \alpha_2 \beta_1 g_{m1} g_{m2}}$ , $C_{eq} = C_A$

The sensitivity coefficients of the simulated equivalent values,  $R_{eq}$ ,  $L_{eq}$  and  $C_{eq}$ , are each affected by the active and passive circuit components, and the finding values are produced as follows.

For parallel R-L;

$$S_{R_B}^{R_{eq}} = 1, S_{\alpha_1, \alpha_2, \beta_1, g_{m1}, g_{m2}}^{L_{eq}} = -1, S_{C_A}^{L_{eq}} = 1. \quad (5)$$

For parallel R-C;

$$S_{\alpha_1, \alpha_2, \beta_1, R_A, g_{m1}, g_{m2}}^{R_{eq}} = -1, S_{C_B}^{C_{eq}} = 1. \quad (6)$$

For parallel L-C;

$$S_{\alpha_1, \alpha_2, \beta_1, g_{m1}, g_{m2}}^{L_{eq}} = -1, S_{C_B}^{L_{eq}} = 1, S_{C_A}^{C_{eq}} = 1. \quad (7)$$

All of the sensitivity coefficients from above (5) to (7) have magnitudes that are less than or equal to one. As a result, the sensitivity of all the proposed parallel R-L, R-C, and L-C immittance simulators is quite low.

#### 4. Simulation Results

PSPICE simulation program has been used to simulate the suggested grounded parallel-type impedance simulator in Figure 3. The simulator was designed employing CMOS VDBA of Figure 2 with a model of 0.18- $\mu\text{m}$  process parameters from TSMC. Table 3 lists the computed aspect ratio (W/L) for each transistor. The supply voltages used to bias this circuit were  $+V = -V = 0.75 \text{ V}$ .

Table 3: Calculated transistor dimensions of VDBA in Figure 2

Transistor	W/L ( $\mu\text{m}/\mu\text{m}$ )
$M_1$ - $M_2$ , $M_5$ , $M_7$ - $M_8$ , $M_{12}$ - $M_{13}$	2.4/0.18
$M_3$ , $M_9$ , $M_{14}$	5/0.18
$M_4$ , $M_{10}$	5.2/0.18
$M_6$	3.25/0.18

$M_{11}$	10/0.18
----------	---------

The following components were chosen for simulations:  $I_{B1} = I_{B2} = 90 \mu\text{A}$  for  $g_m = g_{m1} = g_{m2} = 0.641 \text{ mA/V}$ ,  $R_A = R_B = 1 \text{ k}\Omega$ , and  $C_A = C_B = 50 \text{ pF}$ . Using data from Table 1, the simulated equivalent values of Figure 3 can be derived as:

- for R-L simulator:  $R_{eq} = 1 \text{ k}\Omega$  and  $L_{eq} = 0.12 \text{ mH}$ ;
- for R-C simulator:  $R_{eq} = 2.44 \text{ k}\Omega$  and  $C_{eq} = 50 \text{ pF}$ ;
- for L-C simulator:  $L_{eq} = 0.12 \text{ mH}$  and  $C_{eq} = 50 \text{ pF}$ .

The total power consumed in the circuit for this setting was found to be 0.388 mW.

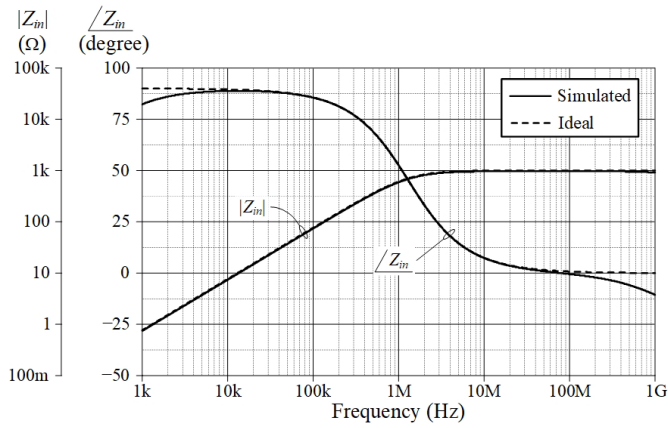
Based on the results of the simulation and theory, Figure 4 depicts the magnitude and phase frequency characteristics of the proposed parallel-type immittance simulator circuit in Figure 3. The frequency corners ( $f_c$ ) of the R-L and R-C impedance simulators in Figs. 4(a) and 4(b) obtained from the simulation results are found to be roughly 1.29 MHz, which is pretty close to the calculated value of 1.30 MHz. In addition, the simulated  $f_c$  value of the L-C impedance simulator was discovered to be 2.04 MHz, which nearly equals to the ideal value of  $f_c = 2.05 \text{ MHz}$ . The input voltage ( $v_{in}$ ) and current ( $i_{in}$ ) responses through the R-C impedance simulator are also displayed in Figure 5 as simulated time-domain waveforms. This performance was evaluated by supplying a sinusoidal input signal with a peak value of 50 mV at  $f = 1 \text{ MHz}$  to the simulated RC impedance circuit.

In order to further illustrate the electronic adjustability of the proposed circuit, the parallel L-C simulator has been performed to change  $I_B = I_{B1} = I_{B2} = 50 \mu\text{A}$ ,  $100 \mu\text{A}$ , and  $200 \mu\text{A}$ , while maintaining  $C_A = C_B = 50 \text{ pF}$ . As a consequence, the simulated equivalent inductance value ( $L_{eq}$ ) has been altered to 0.22 mH, 0.11 mH, and 54.8  $\mu\text{H}$ , respectively, while the simulated equivalent capacitance value ( $C_{eq}$ ) remains constant at 50 pF. The results of the simulated frequency responses compared with the theory are given in Figure 6.

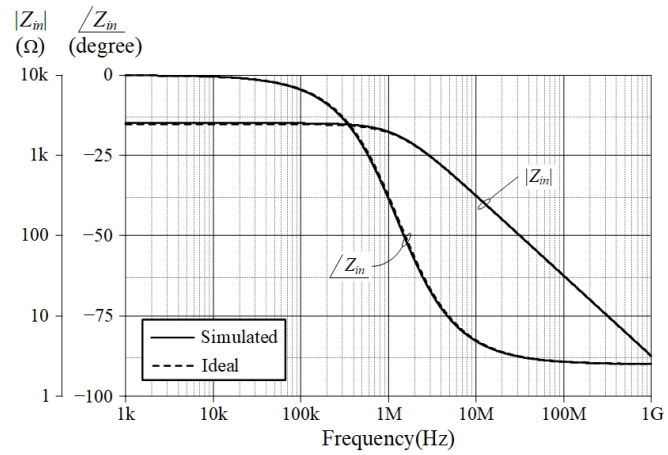
The impact of varying ambient temperature variation on the simulator responses is also being considered. This was accomplished by testing the proposed R-C simulator circuit with changes in ambient temperature ranging from 0°C to 100°C with steps of 25°C. Figure 7 displays the result of its magnitude variations.

#### 5. Experimental Results

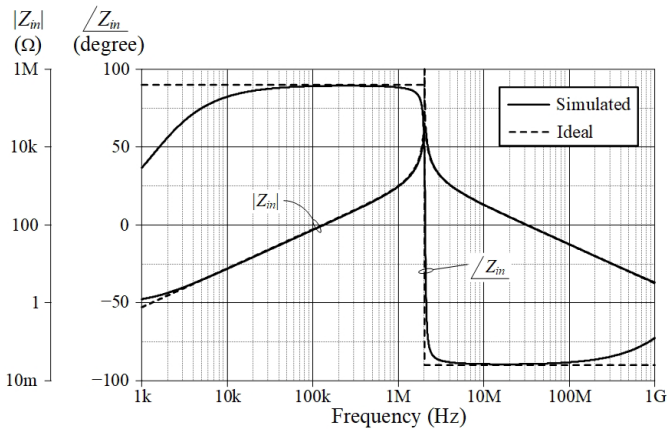
In order to further confirm the feasibility of the proposed idea, the suggested circuit of Figure 3 has been tested in the laboratory utilizing IC LT1228 from Linear Technology [14]. The package information and internal behavior of IC LT1228 are shown in Figure 8. There are two amplifiers: OTA and CFOA. The OTA is used to provide a high-impedance differential input and a current source output with wide output voltage compliance, while the CFOA is utilized to transmit voltage from the z terminal to the o terminal, and the current from the z terminal to the x terminal. According to the following relation, the transconductance gain ( $g_m$ ) of the LT1228 in this case is reliant on the external bias current ( $I_B$ ) [14]:



(a)



(b)



(c)

Figure 4: Ideal and simulated frequency responses of the proposed parallel-type immittance simulator circuit in Figure 3. (a) R-L impedance simulator (b) R-C impedance simulator (c) L-C impedance simulator

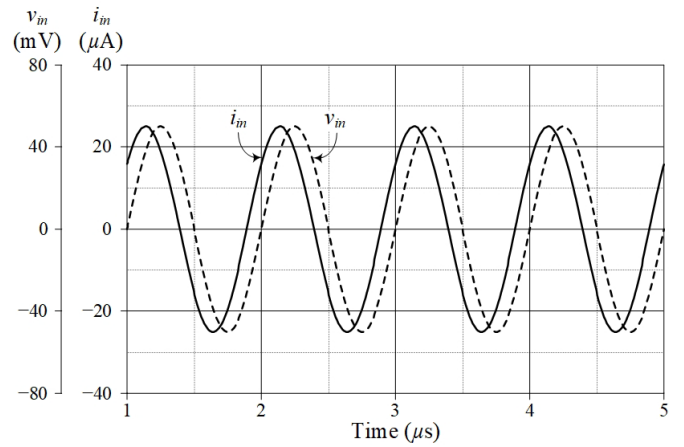


Figure 5: Simulated time-domain responses for  $v_m$  and  $i_m$  of the R-C simulator.

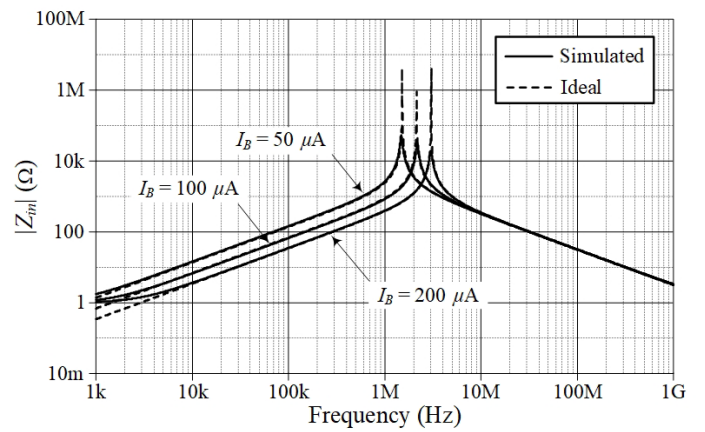


Figure 6: Simulated frequency responses of the L-C simulator with varying  $I_B$ .

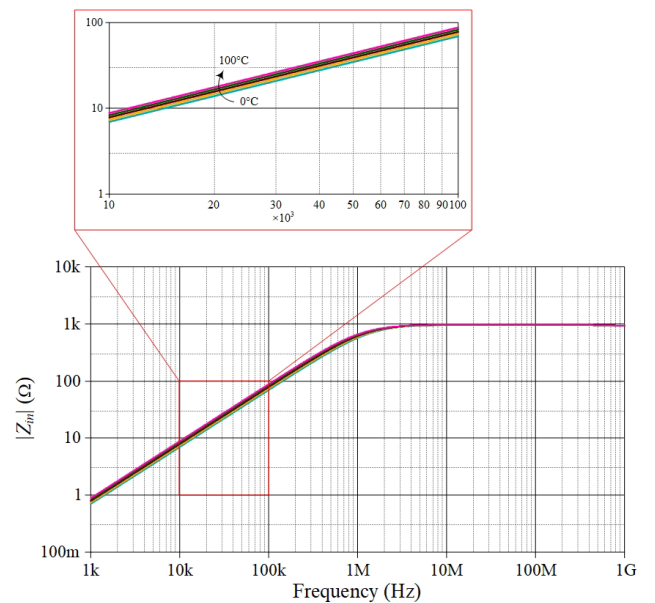


Figure 7: Simulated frequency responses of the R-C simulator at different temperature ( $0^\circ\text{C}$ ,  $25^\circ\text{C}$ ,  $50^\circ\text{C}$ ,  $75^\circ\text{C}$ , and  $100^\circ\text{C}$ ).



$$g_m = 10I_B \quad (8)$$

In the case of parallel R-L impedance simulation, the active and passive components for the experimental measurement were taken as follows:  $g_m = g_{m1} = g_{m2} = 0.5 \text{ mA/V}$  ( $I_B = I_{B1} = I_{B2} = 50 \mu\text{A}$ ),  $R_B = 1 \text{ k}\Omega$ , and  $C_A = 1 \text{ nF}$ , resulting in  $R_{eq} = 1 \text{ k}\Omega$ , and  $L_{eq} = 4 \text{ mH}$ . With symmetrical supply voltages of  $\pm 5 \text{ V}$ , the LT1228 was biased. Figure 9 shows the grounded parallel lossy inductor's measured magnitude and phase responses for the selected components.

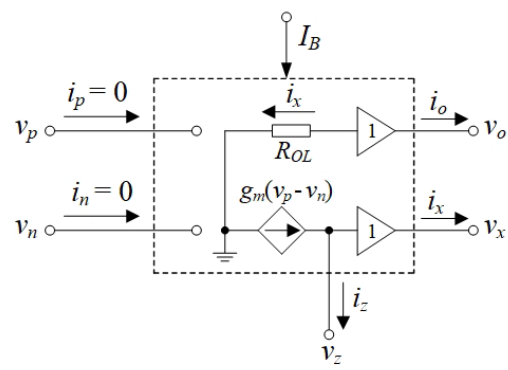
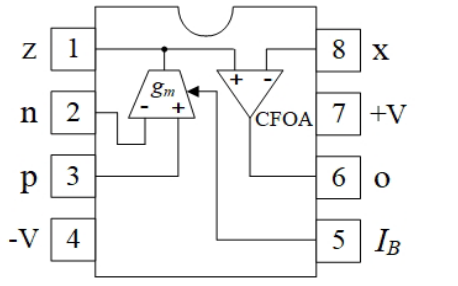
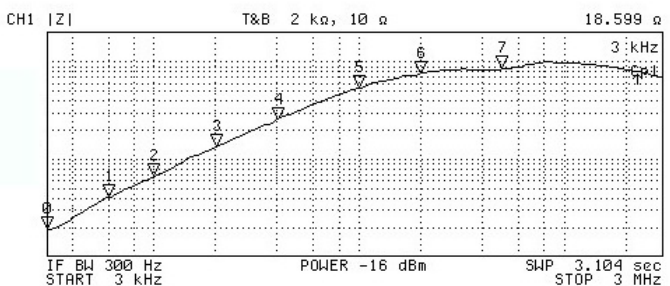
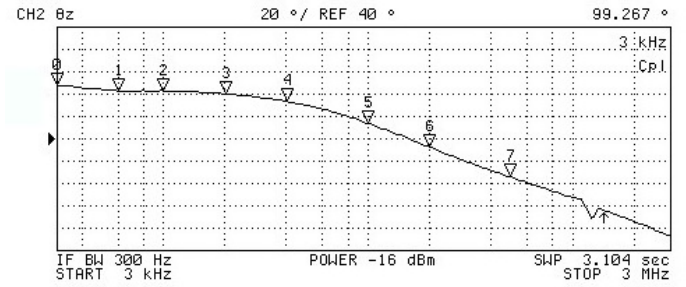


Figure 8: IC LT1228 (a) package information (b) its internal behavior.



N	SWP PARAM	VAL
0	3 kHz	18.599 Ω
1	6 kHz	40.525 Ω
2	10 kHz	65.485 Ω
3	20 kHz	132.23 Ω
4	40 kHz	251.76 Ω
5	100 kHz	536.14 Ω
6	200 kHz	751.61 Ω
7	500 kHz	838.94 Ω

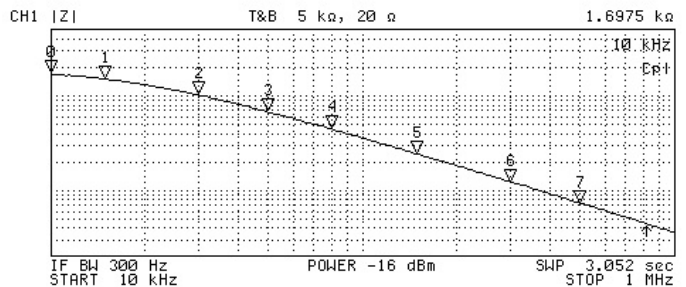
(a)



N	SWP PARAM	VAL
0	3 kHz	94.267 °
1	6 kHz	82.793 °
2	10 kHz	83.121 °
3	20 kHz	80.294 °
4	40 kHz	73.487 °
5	100 kHz	53.942 °
6	200 kHz	32.572 °
7	500 kHz	5.8288 °

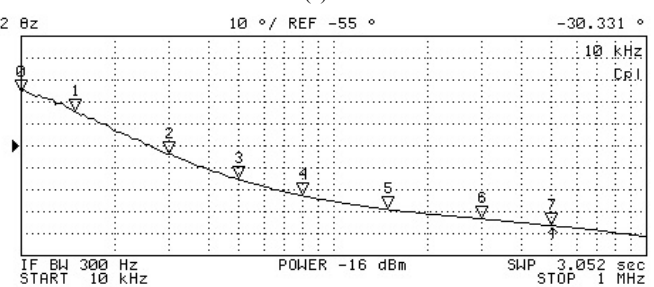
(b)

Figure 9: Measured frequency responses of the parallel R-L simulator (a) magnitude response (b) phase response



N	SWP PARAM	VAL
0	10 kHz	-30.331 °
1	15 kHz	-39.565 °
2	30 kHz	-58.712 °
3	50 kHz	-70.298 °
4	80 kHz	-77.599 °
5	150 kHz	-83.904 °
6	300 kHz	-88.317 °
7	500 kHz	-91.252 °

(a)



N	SWP PARAM	VAL
0	10 kHz	-30.331 °
1	15 kHz	-39.565 °
2	30 kHz	-58.712 °
3	50 kHz	-70.298 °
4	80 kHz	-77.599 °
5	150 kHz	-83.904 °
6	300 kHz	-88.317 °
7	500 kHz	-91.252 °

(b)

Figure 10: Measured frequency responses of the parallel R-C simulator (a) magnitude response (b) phase response

The parallel R-C simulator was then tested with the following parameters:  $g_m = g_{m1} = g_{m2} = 1 \text{ mA/V}$  ( $I_B = I_{B1} = I_{B2} = 100 \text{ }\mu\text{A}$ ),  $R_A = 500 \text{ }\Omega$ , and  $C_B = 4.7 \text{ nF}$ , yielding  $R_{eq} = 2 \text{ k}\Omega$ , and  $C_{eq} = 4.7 \text{ nF}$ . Figure 10 shows the measured frequency responses for the equivalent input impedance of the simulator. The experimental results shown in Figures 9 and 10 demonstrate the suggested circuit's practicality in application areas.

### 6. Application Example

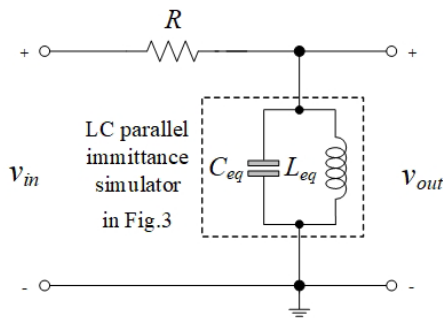
The second-order voltage-mode bandpass filter in Figure 11(a) is intended to emphasize operational performance as an illustration of an application. The bandpass filter realization utilizing the proposed L-C parallel immittance simulator in Figure 3 is shown in Figure 11(b). The voltage transfer action of the filter is written as:

$$\frac{V_{out}(s)}{V_{in}(s)} = \frac{s \left( \frac{1}{RC_{eq}} \right)}{s^2 + s \left( \frac{1}{RC_{eq}} \right) + \frac{1}{L_{eq}C_{eq}}} \quad (9)$$

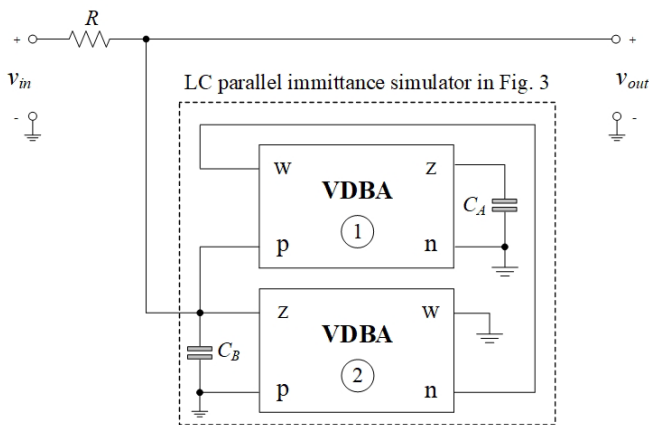
The natural angular frequency ( $\omega_o$ ) and the quality factor ( $Q$ ) of the filter in Figure 11 are determined from (9), respectively, by:

$$\omega_o = 2\pi f_o = \sqrt{\frac{1}{L_{eq}C_{eq}}} \quad (10)$$

and 
$$Q = R \sqrt{\frac{C_{eq}}{L_{eq}}} \quad (11)$$



(a)



(b)

Figure 11: Second-order voltage-mode bandpass filter (a) prototype passive structure (b) utilizing the L-C simulator in Figure 3.

The simulated frequency response of the implemented active bandpass filter is demonstrated in Figure 12 with the following components:  $R = 1.5 \text{ k}\Omega$ ,  $g_m = g_{m1} = g_{m2} = 0.675 \text{ mA/V}$  ( $I_B = I_{B1} = I_{B2} = 100 \text{ }\mu\text{A}$ ), and  $C_A = C_B = 100 \text{ pF}$ . With  $L_{eq} = 0.22 \text{ mH}$  and  $C_{eq} = 100 \text{ pF}$ , the filter is designed to obtain  $f_o = \omega_o/2\pi = 1.07 \text{ MHz}$  and  $Q = 1$ . The resulting responses demonstrate that the circuit can operate correctly between 500 kHz and 100 MHz.

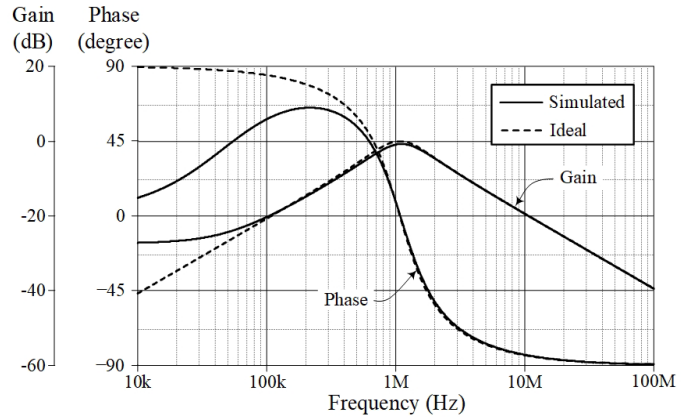


Figure 12: Ideal and simulated results of the bandpass frequency responses in Figure 11(b).

### 7. Conclusions

A grounded parallel RL, RC, and LC impedance simulator has been designed with VDAs and two grounded passive components. Through the use of the transconductance parameter in VDA, the simulated equivalent values, i.e.,  $R_{eq}$ ,  $L_{eq}$ , and  $C_{eq}$ , can all be electronically altered. The circuit has been simulated using the PSPICE program, which is based on 0.18- $\mu\text{m}$  CMOS technology, to demonstrate its viability. In-depth laboratory tests have also been conducted to verify the practical usability of the simulator circuit. The design of a second-order voltage-mode bandpass filter using the proposed simulator is given.

### Conflict of Interest

The authors declare no conflict of interest.

### Acknowledgments

This work was supported by Rajamangala University of Technology Rattanakosin (RMUTR). The support by the School of Engineering, King Mongkut's Institute of Technology Ladkrabang, contact no. 2562-02-01-004, is also gratefully acknowledged.

### References

- [1] D. Biolk, R. Senani, V. Biolkova and Z. Kolka, "Active elements for analog signal processing: classification, review, and new proposals", *Radioengineering*, **17**(4), 15–32, 2008.
- [2] R. Sotner, J. Jerabek, N. Herencsar, N. "Voltage differencing buffered/inverted amplifiers and their applications for signal generation", *Radioengineering*, **22**(2), 490-504, 2013.
- [3] W. Tangsrirat, "Actively Floating lossy inductance simulators using voltage differencing buffered amplifiers," *IETE Journal of Research*, **65**(4), 446–459, 2018, doi.org/10.1080/03772063.2018.1433082
- [4] F. Kaçar, A. Yeşil, A. Noori, "New CMOS realization of voltage differencing buffered amplifier and its biquad filter applications," *Radioengineering*, **21**(1), 333–339, 2012.

- [5] N. Roongmuanpha, T. Pukkalanun, W. Tangsrirat, "Practical realization of electronically adjustable universal filter using commercially available IC-based VDBA," *Engineering Review*, 41(3), 1-14, 2021, doi.org/10.30765/re.1547.
- [6] M. Faseehuddin, N. Herencsar, S. Shireen, W. Tangsrirat, S. H. M. Ali, "Voltage differencing buffered amplifier-based novel truly mixed-Mode biquadratic universal filter with versatile input/output features", *Applied Sciences*, 12(3), 2022, doi.org/10.3390/app12031229.
- [7] P. Moonmuang, T. Pukkalanun, W. Tangsrirat, "Floating/grounded series/parallel R-L, R-C and L-C immittance simulators employing VDTAs and only two grounded passive elements," *AEU - International Journal of Electronics and Communications*, 145, 154095, 2022, doi.org/10.1016/j.aeue.2021.154095.
- [8] P. Moonmuang, W. Tangsrirat, "Single VDTA-based tunable floating lossy inductance simulation circuits," *IETE Journal of Research*, doi: 10.1080/03772063.2021.1900752.
- [9] N. Roongmuanpha, W. Tangsrirat, "Practical floating capacitance multiplier implementation with LT1228s," *Informacije MIDEM- Journal of Microelectronics, Electronic Components and Materials*, 51(1), 85-94, 2021, doi.org/10.33180/InfMIDEM2021.106
- [10] A. Paul and D. Patranabis, "Active simulation of grounded inductors using a single current conveyor", *IEEE Transactions on Circuits and Systems*, 28(2), 164-165, 1981, doi.org/10.1109/TCS.1981.1084947
- [11] H. Kuntman, M. Gülsoy and O. Çiçekoğlu, "Actively simulated grounded lossy inductors using third generation current conveyors", *Microelectronics Journal*, 31(4), 245-250, 2000, doi.org/10.1016/S0026-2692(99)00108-1
- [12] O. Çiçekoğlu, A. Toker and H. Kuntman, "Universal immittance function simulators using current conveyors", *Computers and Electrical Engineering*, 27(3), 227-238, 2001, doi.org/10.1016/S0045-7906(00)00018-5
- [13] F. Kaçar and H. Kuntman, "CFOA-based lossless and lossy inductance simulators", *Radioengineering*, 20(3), 627-631, 2011.
- [14] Linear Technology, "100MHz current feedback amplifier with DC gain control", LT1228 datasheet, 1994.

## A Model for Teaching Mathematics to Gifted Students Based on an Effective Combination of Various Approaches for their Preparation

Zhanna Dedovets<sup>\*1</sup>, Mikhail Rodionov<sup>2</sup>, Anna Novichkova<sup>3</sup>

<sup>1</sup>Department of School of Education, The University of the West Indies (UWI), Trinidad and Tobago

<sup>2</sup>Department of Computer and Mathematical Education, State Pedagogical University named after S. Ayni (TSPU), Tajikistan

<sup>3</sup>Gymnasium named after Abulfazli Balami, Vahdat, Tajikistan

### ARTICLE INFO

Article history:

Received: 27 November, 2022

Accepted: 15 January, 2023

Online: 24 February, 2023

Keywords:

Mathematically gifted students

Model of teaching mathematics

Mathematical Olympiads

Discovery-based method

Partial discovery-based method

Problem-based learning

Project-based learning

### ABSTRACT

Currently one of the urgent goals of mathematical education is the organization of effective work with gifted students. Based on the study of various approaches to teaching mathematically gifted students, many years of experience of teachers, students' work, and an analysis of curricula and materials for schools with in-depth study of mathematics, an author's model for the training of gifted students was developed. The novelty of this model is that it ensures a rational combination of various forms of education for gifted children on the basis of differentiation, individualization of the process of teaching mathematics, advanced learning, openness, democracy, reflection, and adequate control. The pedagogical experiment was carried out for two years in the Abulfazl Balami gymnasium for gifted children in the city of Vahdat, Republic of Tajikistan. 41 students and 18 teachers took part in the experiment. The data obtained from the experimental and control groups were subjected to qualitative and quantitative analysis. Over the same time interval there were significant changes in the performance of students in the experimental groups, with 40% of the students moving to a higher level. In the control groups, the change was not significant.

## 1. Introduction

As is well-known, modern society sets complex tasks for citizens that require non-standard decisions and the manifestation of critical and creative thinking in the constantly changing and non-predictable environment of our world. Accordingly, in a modern school, the discovery-based method as an effective aid for working with gifted children and its optimal organization, combining various formats of such work, is increasingly coming to the fore [1-3].

Despite many studies of the theory and practice of working with gifted children, this issue has not been sufficiently discussed in the methodological and mathematical studies known to us, which indicates the relevance and importance of the present study.

The response to this challenge should consider the specifics of the subject. The process of teaching mathematics to gifted students at school seems to be quite laborious, since here it is necessary, firstly, to ensure the development of the main curriculum, and, secondly, to effectively develop their "non-standard potential".

These two goals in existing educational practice are not always coordinated. Elective and basic mathematics courses are taught by different teachers under time pressure. Moreover, this often do not take into account the individual characteristics of students.

As a rule, in the real educational process, the work of a mathematics teacher, both in ordinary and specialized classes, focuses primarily on the development of students' basic competencies, which are taken into account during the current and final certification. The second component of this process - the actual mathematical development - is decided on elective courses and consultations of various kinds by completely different people - "invited lecturers" within their areas of expertise. These two groups of teachers do not always have the opportunity and desire to closely contact each other professionally. As a result, gifted students study, as it were, two different subjects: "basic mathematics" and "Olympiad mathematics", which does not always allow them to effectively realize the developing potential of the studied mathematical content. This was confirmed by our survey of mathematics teachers, most of whom indicated that the spontaneous interaction of the various formats of such training, as

\*Corresponding Author: Zhanna Dedovets [dedzhanna333@gmail.com](mailto:dedzhanna333@gmail.com)

occurs in the existing system of training gifted children, does not fully ensure its effectiveness. From the foregoing, it is evident that there is a need for theoretical development and practical implementation of a special model for the rational combination of basic and elective courses in the process of teaching mathematics to gifted students. Summarizing, we can conclude that as a rule, the traditional strategy for teaching mathematics to gifted children is one-sided, being limited to an emphasis on elective courses, or individual consultations. At the same time, the learning process itself is spontaneous in nature, not providing for the educational needs of all such students. An analysis of the pedagogical studies known to us revealed that the issue of a rational combination of various formats for mathematical training of gifted schoolchildren was not specifically addressed in these studies.

Thus, the relevance of this study stems from the need for theoretical development and practical implementation of a holistic strategy for working with gifted children in the field of mathematics, which includes the possibility of a rational combination of various learning formats for each gifted student.

## 2. Mathematically Gifted Students

Many psychological and pedagogical studies have been devoted to the mathematical development of gifted children [4-10]. According to most authors, mathematical giftedness is understood as a kind of intellectual giftedness, which is associated with and develops in special mathematical activity. Its basic characteristics are integrity, multicomponent nature, hierarchy and dynamism [11, 12]. Mathematically gifted schoolchildren are characterized by the ability to think logically, the ability to operate with mathematical symbols, quickly and correctly solve mathematical problems, successfully moving from simple to more complex mathematical constructions.

Such students have a flexible mind, that is, they are able to find a way out of a non-standard mathematical situation and they have a well-developed abstract memory [13]. Approaches to the study of mathematical giftedness, reflected in the literature, are very diverse: they are based on the psychology of individual differences in students, on the special abilities of mathematically gifted students.

The scientists Joy Gilford, Ellis Torrance, Frank Barron and Charles Taylor carried out a number of major studies in the psychology of giftedness and contributed to the unification of theoretical studies on the psychology of individual differences and practical work on the construction of new curricula in the field of differentiated learning [14-18]. They found that giftedness was manifested in the fact that, unlike for ordinary, traditional experiments, students built their own tasks. Scientists have observed the behavior of creatively gifted people in natural situations of communication, work and leisure. They tried to determine the specific manifestations of talent in various activities, as well as the characteristic features of the personality of gifted people, which emerged in behavior, thinking, inclinations and attitudes. The tasks were set to change the idea of giftedness as "a symptom of hereditary degeneration of the epileptoid type" [19]. This process of change took place over a period of more than 30 years from the beginning of the 20th century. The results of their research have shown that by the end of school, many gifted children sometimes experience severe depression. They are forced

to hide their giftedness from their peers and adults. Gifted children experience "discrimination" in school due to the lack of differentiated teaching, due to the school's focus on the average student, due to excessive reduction to a uniform system of curricula [20].

Psychologists Sergei Rubinstein and Boris Teplov developed a classification of the concepts of "ability", "giftedness" and "talent". The classification was carried out according to the success of the activity [16]. Abilities are considered as individual psychological characteristics that distinguish one person from another, on which the possibility of success in activity depends, and giftedness is considered as a qualitatively unique combination of abilities (individual psychological characteristics), on which the possibility of success in activity also depends.

In various definitions of the concept of giftedness (source), a number of basic features of giftedness can be traced. A person has:

- (1) outstanding (high level) abilities,
- (2) developed intelligence,
- (3) an increased level of mental development,
- (4) creative approach,
- (5) the possibility of achieving high results in various activities.

Intellectual giftedness is a developing systemic quality of the personality psyche in the structure of general abilities. The development of this quality requires a holistic didactic approach to working with gifted adolescents [6]. The personal growth of intellectually gifted adolescents depends on the type of educational environment. The environment should contribute to the disclosure and optimal manifestation of the creative nature of the psyche of gifted adolescents. By minimizing the difficulties of a gifted child in contact with his environment, the educational environment contributes to the adequate personal development of gifted adolescents [21].

Professor Gennadiy Sarantsev in his works talks about methods for developing the ability of gifted children to solve non-standard tasks. He considered various heuristic approaches to solving problems and building new curricula in the field of differentiated learning.

Scientists Vadim Krutetsky, Victoria Yurkevich, Irina Levochkina, Elena Kryukova examined in detail the special abilities that characterize mathematically gifted students, as well as ways to recognize them in a child and adolescent.

They emphasize that schoolchildren who are especially gifted in mathematics are characterized by a peculiar mathematical orientation of the mind (the tendency to perceive many phenomena through the prism of mathematical relations, to realize them in terms of logical and mathematical categories).

They single out several components of mathematical talent: the ability to arrive at mathematical generalization; rationality of the decision (the ability to find the shortest way to solve, cut off the excess, not directly related to the achievement of the goal, the task); a sense of "mathematical aesthetics" (the ability to see beauty and elegance in a simple and at the same time witty, concise and economical way of solving a problem); reversibility of thinking; mathematical intuition.

The results of studies of mathematically gifted adolescents show that adolescents gifted in mathematics develop such features of their mental activity as the ability to generalize mathematical material (the ability to see the general in what is outwardly different, singular), the flexibility of thought processes and the desire to find simpler but more effective ways to solve problems [22].

The issue of teaching methods for mathematically gifted schoolchildren is presented in the works of many psychologists and teachers [11, 23]. They consider various methods of working with mathematically gifted students, describe a system of special tasks for gifted students, and describe the necessary conditions for creating a support system for talented students. They discuss the technologies that the teacher should rely on when developing the unique abilities of each child, describe the construction of a program for mathematically gifted students, where they propose to integrate Olympiad tasks into sections of the basic mathematics course (resources). A number of authors describe the development of creative potential in the process of participating in competitions, in the process of solving non-standard problems [24, 25]. They view work with gifted children in the context of differentiated and individualized learning. Such training, as is well known, is implemented on the basis of a full account of the individual and typological characteristics of a person in the form of grouping or ungrouping students and differentiated construction of the learning process in educated groups or individually; learning technology, the purpose of which is to create conditions for the identification of existing inclinations, the effective development of the interests and abilities of students [26].

However, the authors known to us do not specifically consider the correlation and connections of various approaches to differentiated work with gifted students, the conditions for their effective integration within the framework of such work, the difficulties that arise in this, and ways to overcome them. The foregoing determines the relevance and significance of building and creating a model for training gifted students in mathematics, based on a rational combination of various approaches for working with them and, in particular, in both basic and elective mathematical courses.

In the context of our study, the "Working Concept of Giftedness" developed by Diana Bogoyavlenskaya and Vladimir Shadrikov is of great interest [7]. This concept involves the disclosure of giftedness on the basis of the theoretical provisions of psychology and the definition of basic principles in solving the problems of identifying, training and developing gifted children. 'Giftedness' is interpreted as a systemic quality that characterizes the child's psyche as a whole. At the same time, it is the personality, its orientation and the system of values that lead to the development of abilities and determine how its potential will be realized.

Giftedness entails a humanistic approach to the education and development of gifted students, that is, special attention is paid to caring for a gifted child, which implies an understanding of not only the advantages, but also the difficulties that his giftedness brings with it.

### 3. Methodological framework

As a basis for building the authors' model of work with gifted schoolchildren in mathematics, they used differentiated and individual approaches to learning and the above-mentioned "Working concept of giftedness" [15, 27]. Differentiation of learning is a process involving the division of students into groups according to their mathematical abilities. Individualization of learning - learning aimed at developing the individual abilities of each student - is an integral element of student-centered learning. Such work may include, for example, external studies, elective courses, individual consultations, implementation of project activities. At the same time, it should not lead to the need to separate gifted students from their peers, but should involve the integration of various collective, group and individual learning activities.

As you know, today, work with gifted children is carried out in schools, gymnasiums, and lyceums. In particular, differentiated education is carried out by dividing classes into profiles: physical and mathematical, humanitarian, chemical and biological, and others. At the same time, work with gifted children is carried out both within the framework of school lessons and within the framework of additional courses (elective courses, courses for preparing for Olympiads, individual lessons). A rational combination of these formats makes it possible to implement regular and systematic work to maximize the full disclosure of the creative potential of schoolchildren [26, 28].

To organize such work, the following factors are necessary:

- (1) a strategy for teaching gifted students heuristic methods for solving problems,
- (2) continuity of basic and elective mathematical courses of a developing orientation,
- (3) a system of students-centered methods of working with gifted students.

### 4. A strategy for teaching gifted students' heuristic methods for solving non-standard problems

The main forming factor in organizing such work is the strategy of teaching gifted schoolchildren heuristic methods for solving non-standard problems. Non-standard problems are understood as problems that cannot be solved by standard algorithms known from the basic mathematical course. When solving them, it is necessary to use one or another heuristic procedure.

The essence of heuristic methods for solving problems lies in the fact that the student is naturally involved in the process of rediscovering a non-standard condition of the problem and finding a way to solve it, without having a direct opportunity to apply the basic algorithm for solving. The selection of such tasks and the development of an individual learning plan is a serious challenge for a mathematics teacher. This plan should include the possibility of targeted implementation of several individual learning approaches that correspond to different strategies for working with a particular gifted student.

Scholars define an educational plan as a set of learning stages, forms of learning, and combinations of individual topics from

mathematics curricula. In other words, the educational plan is a differentiated educational program that provides the student with a choice of the type and amount of pedagogical support he needs for his self-determination and self-realization [28].

An individual educational plan is built based on the educational needs, individual abilities and capabilities of the student (level of readiness to master the program), as well as existing regulatory documents. The purpose of such training is to purposefully ensure the differentiation and individualization of the education of gifted children, giving it a personality-oriented character.

An individual educational plan consists of a number of “sections” corresponding to various formats of work (basic course, elective course, individual work, group projects, competitions, Olympiads, etc.). All these approaches should be closely related and organically combined with each other.

**5. Ensuring the continuity of the basic and elective components of the mathematical teaching of gifted children**

In the combination of various formats of work with gifted children in mathematics mentioned above, a special role is played by the continuity of basic and elective mathematical courses of a developing orientation. This factor, as the analysis of the literature and our own pedagogical experience show, has not yet become one of the imperatives of the educational process for the considered contingent of schoolchildren.

In particular, often different mathematical courses in the same group are taught by teachers of different qualifications (school and university), while the material studied in parallel within these courses is often characterized by "diversity", "patchwork" both in content and in developmental aspects. Overcoming such diversity, obviously, involves providing an organic combination of basic (profile) and elective courses. To illustrate the latter point, we present Figure 1, in which the first and third columns present some topics that are quite important in terms of preparing schoolchildren for mathematical Olympiads. In the central part of the diagram, sections are presented that are the result of the “interaction” of the corresponding basic and additional mathematical courses (Figure 1).

Commenting on the structure and content of the above diagram, it is necessary, first of all, to note that work with gifted students is a holistic, systematic process that emerges from the basic course.

Here, when solving developmental problems, students mainly apply the heuristic method at the stage when they discover some general algorithms that are practiced with all students in a collective form. On the elective course, in a differentiated or individual form, the studied material is deepened by varying and combining the mastered algorithms when solving problems of a search nature. Such work, in turn, creates the basis for mastering various heuristic procedures by gifted students, which, in particular, further contributes to their successful participation in Olympiads at various levels.

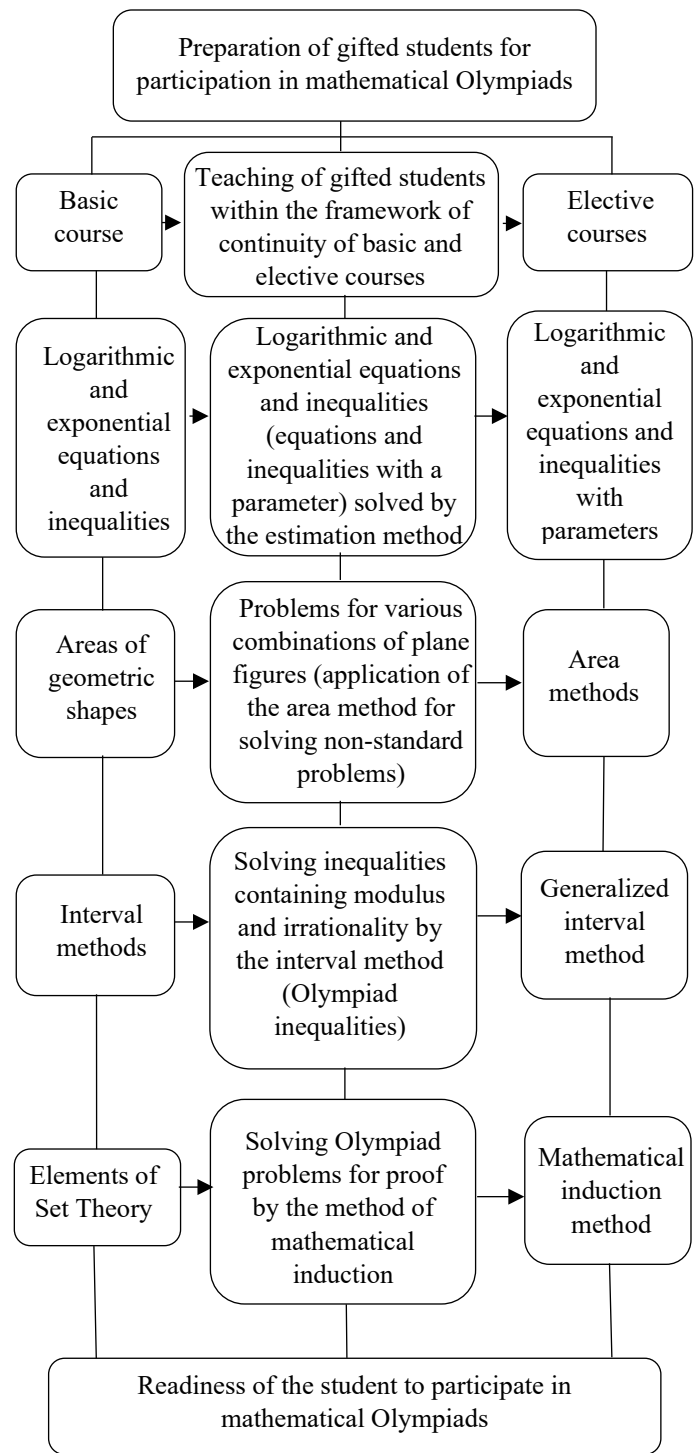


Figure 1: Continuity of Basic and Elective Courses in the Preparation of Students in Grades 4-5 for Olympiads in Mathematics

For example, within the framework of the basic course, the traditional topic "Method of intervals" is studied. Here we touch upon square inequalities, cubic inequalities, inequalities of higher degrees, fractional rational inequalities solved by the interval method based on the sign placement rule. Within the framework of the elective course "Selected Issues of Mathematics", it is advisable to consider the generalized method of intervals, which is used, in particular, in solving irrational and trigonometric inequalities, as well as inequalities containing unknowns under the

module sign, unknowns in the exponent, inequalities containing logarithms of the type:

$$\left(x + \frac{5}{x}\right) \left(\frac{\sqrt{x^2-10x+25-1}}{\sqrt{6-x-1}}\right)^2 \geq 6 \left(\frac{\sqrt{x^2-10x+25-1}}{\sqrt{6-x-1}}\right)^2.$$

The consideration of such inequalities is directly based on the material of the basic course, while providing for the variable application of the known algorithm.

At the same time, tasks of a search (Olympiad) nature are beginning to be involved, which are an example of tasks of an even more generalized nature.

Problem: For what values of the parameter  $a$  among the solutions of the inequality  $(x^2 - ax - x + a)\sqrt{x + 5} \leq 0$ , will there be two solutions, the difference between which is equal to 4?

When solving this problem, obviously, the generalized method of intervals is used and the analysis of all possible cases of the state of the considered mathematical construction is carried out, depending on the value of the parameter  $a$ . Accordingly, when analyzing a possible solution path, it is advisable for the teacher, if necessary, to rely on the material of the basic mathematics course, projecting it onto a higher level of generalization of the content.

Irrational and trigonometric inequalities are included in the curriculum for mathematics in high school. Here they are considered to be inequalities, the solution of which, depending on their complexity, is carried out by the method of intervals both at the basic and advanced levels (in the elective course), as well as in preparation for the Olympiads. The method of intervals in various modifications is used at all stages of the study of inequalities, providing a connection, in the context under consideration, between the relevant substantive sections of the basic and elective courses.

The interval method can be used in solving irrational inequalities of a certain type, subject to the appropriate restrictions arising from the properties of the arithmetic root of the  $n$ th degree. The algorithm for using this universal method is well known.

### 6. Model of preparation of gifted children in mathematics

Based on the idea of purposefully ensuring the continuity of basic and elective courses, we have built and implemented a methodological model for working with gifted students in mathematics (Table 1).

Table 1: Model of Preparation of Gifted Children in Mathematics

Purpose: Creation of conditions for the development of mathematical abilities of gifted students, their self-development, the harmonious development of the personality of a unique child; ability to independently acquire and apply knowledge	
Tasks:	<ol style="list-style-type: none"> <li>1. Identification of gifted students in the field of mathematics</li> <li>2. Development of methodological support for the effective development of the mathematical abilities of gifted students, providing for the continuity of various forms of their preparation</li> <li>3. Implementation of mathematical training of schoolchildren, aimed at achieving socially and personally significant results</li> </ol>

4. Approbation and monitoring of the proposed methodological solutions	
Principles of building the educational process of gifted students	
The principle of rational combination of various forms of work with gifted children, the principle of differentiation and individualization of the learning process, the principle of student-centered learning, the principle of advanced learning, the principle of openness, the principle of adequate control, the principle of democracy, the principle of reflection	
The content of work with gifted children in mathematics	Profile course When teaching gifted children in mathematics, the existing programs of specialized courses and relevant textbooks are involved
	Elective course When teaching children gifted in mathematics, topics are considered that deepen the relevant topics of the profile course, and topics that go beyond its scope (for example, the method of mathematical induction, graph theory, etc.)
	Project work When working on an individual project, attention is paid to topics that go beyond the core and basic courses, topics that affect the relationship of mathematics with other areas of knowledge, non-standard solutions to standard tasks are considered (for example, the project "Ten ways to solve one quadratic equation")
	Preparation for the Olympics When preparing schoolchildren for participation in the Olympiads, first of all, various sets of tasks are considered, which include non-standard tasks (e.g. coloring problems, double counting problems)
Forms and methods of organizing work with gifted children in mathematics	
Forms of work organization 1. Standard school lesson as part of the basic course 2. Standard school lesson within the profile course 3. Elective courses in mathematics 4. Math events 5. Work on an individual project 6. Individual preparation for the Olympiads 7. As a result of work on all previous formats - participation in subject Olympiad	Dominant methods of organizing work 1. Discovery-based method 2. Partial discovery-based method 3. Problem-Based learning 4. Project-Based Learning
Expected learning outcomes for gifted students	
<ol style="list-style-type: none"> <li>1. Successful participation of students in Olympiads and conferences</li> <li>2. Readiness to pass exams</li> <li>3. The formed ability of students to independently acquire and apply knowledge in the framework of project-based learning</li> </ol>	



4. Increasing motivation to work on solving problem of a discovery-based method and partial discovery-based method nature
Criteria for success in working with gifted students
It is necessary to understand how much the student's giftedness was enhanced, which is manifested in the following indicators of the student's preparation:
<p>1) can work with mathematical text, solve text problems of increased difficulty</p> <p>2) can solve non-standard problems using heuristic methods</p> <p>3) can solve problems of an increased level of complexity by various methods, choosing from them the more rational one</p> <p>4) can effectively solve problems of an Olympiad nature</p> <p>5) has high-level thinking according to Bloom's taxonomy, has high levels of cognitive ability:</p> <p>a) identifies hidden (implicit) assumptions in the problem, evaluates the significance of the data (analysis)</p> <p>b) uses knowledge from various fields to make a plan for solving a non-standard problem (synthesis)</p> <p>c) has the ability to evaluate particular mathematical material, that is, he can select and study in-depth material based on different criteria (evaluation)</p>

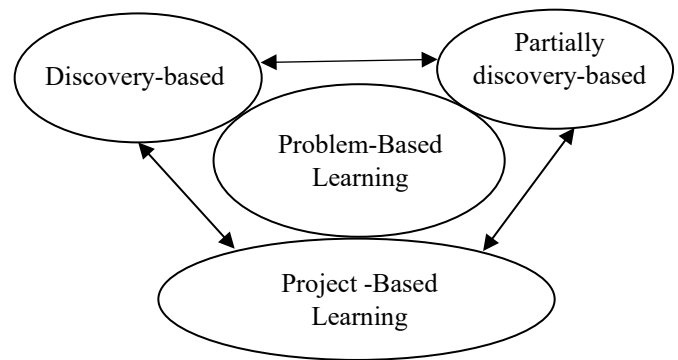


Figure 2: The relationship of various methods of working with gifted students

### 6.1. Partially discovery-based teaching method

In order to gradually bring students closer to independent problem solving, including non-standard tasks, it is necessary to first teach how to perform individual steps of solving a problem, individual stages of search [1, 2, 6, 29]. In this case, the partial discovery-based method of teaching should be used. It is necessary to offer students independent work with a mathematical text, to provide them with the opportunity to independently derive and formulate the basic mathematical concepts by means of a plan devised using heuristic conversation, well-known sources and their own search work. Another variant of this method is to break down a complex problem into a series of available subproblems, each of which makes it easier to approach the solution of the main problem. To complete a separate part of the task, a gifted student has to draw on knowledge from various branches of mathematics, which he could meet both when studying the basic program, and in elective classes. Thus, within the framework of this method, the teacher constructs a task, divides it into auxiliary ones, outlines the steps to find a solution to the problem. Students, on the other hand, perceive the task, comprehend its condition and solve part of the problem, actualizing the available knowledge and discovering the necessary information, exercising self-control in the process of arriving at a solution and self-motivating. But at the same time, the student's activity does not involve planning the research stages and correlating the stages with each other. The partial discovery-based method is used within the framework of the basic course and elective classes.

### 6.2. Discovery-based teaching method

The next method that should be introduced into the process of teaching gifted students is the discovery-based method of teaching. Within the framework of this method, the activities of students target the independent resolution of practical problems that require creative solutions—hypotheses [1, 13, 30]. The teacher gradually moves away from drawing up a plan for solving problems, dividing them into subtasks, and invites gifted students to put forward their own hypotheses to find a solution. This method can be used both during the development of a profile course when solving problems of an increased level of complexity (for example, problems with parameters), and in the process of working on a student research project. This method is actively

As can be seen from this table, the formulation of this model is guided by the principles of a rational combination of various forms of work with gifted children, differentiation and individualization of the learning process, student-centered learning, advanced learning, openness, democracy, reflection and adequate control.

Let us briefly explain the content of the last three principles in our understanding (the content of the rest is obvious).

The principle of openness implies an approach to the subject of study as potentially open, allowing constant expansion and generalization by connecting initially non-obvious meaningful relationships.

The principle of democracy presupposes the right of the student to voluntarily choose the level and the corresponding form of education that he considers most acceptable.

The principle of reflection implies the need for the teacher to constantly monitor the nature of the interaction of a gifted student with peers, his behavior in situations of success and failure. In the course of devising a problem, it is necessary to carry out an ongoing adjustment of the individual plan for the training and education of a gifted student.

Finally, the principle of adequate control presupposes a variety of diagnostic tools for assessing the mathematical training and development of gifted schoolchildren, which are not limited to existing regulatory documents (control and independent work, exam materials, competitions and Olympiads of various levels).

This diagram shows the dominant methods of working with gifted students, their relationship. We will reveal the essence of each of them. All these methods are based on the active discovery-based and creative activity of students.

used in solving Olympiad problems. Initially, the student expresses a hypothesis for solving such a problem, and then builds his own small study to prove or disprove it. Also, this method will be appropriate when mastering new knowledge. An important role is played by independent experiments on the derivation of basic mathematical concepts and statements.

### 6.3. Problem-based Learning

Within the framework of the data of the problem-based method of teaching, a gifted student always faces a problem that requires a creative, heuristic approach to its solution [25, 31-33]. First, in the basic course, at the very beginning of the elective courses, the teacher himself puts forward small problem situations to the students, dividing the more complex ones. Further, the students themselves meet and recognize them, organizing their research activities to find solutions to these problem situations. With the help of the problem-based method of teaching, the skill of independent search work is formed, which helps in preparing for the Olympiads, when working on an individual research project. The problem-based method is also implemented in the process of participation by schoolchildren in various mathematical festivals and mathematical Olympiads. In these formats, a gifted student is constantly faced with a new, partially or completely unknown mathematical situation that requires a heuristic solution.

### 6.4. Project-based Learning

One of the most difficult methods for organizing the activities of schoolchildren is the project method. The essence of the project method is the solution of a problem based on the independent activity of students using appropriate methods, means, knowledge, including interdisciplinary, intellectual and practical skills, as well as the realization of creative potential to obtain a specific result [27, 34-35, 39].

In our opinion, the use of projects as applied to gifted students, first of all, helps to maintain constant motivation for an in-depth study of mathematics. Students can choose an interesting topic for themselves, for example, explore different ways to solve one problem and study the proof of a little-known theorem. Thus, the student is constantly developing independently and expanding his knowledge with interesting mathematical facts primarily for himself [36-38].

The project method can also be used as part of a school lesson, an elective course, or an extracurricular activity. It is possible to offer schoolchildren the opportunity to independently acquire knowledge in solving practical problems and problems that require the integration of knowledge from various subject areas. Similar problems are encountered in preparing for examinations in mathematics.

Revealing the structural elements of Figure 2 and their relationship, we can conclude that the proposed system of methods is aimed at developing the research abilities of gifted students and their creative potential in solving non-standard

problems. Each method can be implemented or partially implemented in different learning formats. That is, this system of methods is the main integration factor for various formats of training gifted students. Using a system of methods, it is possible to construct a diagram of the relationship between various teaching formats in the preparation of mathematically gifted students (Figure 3).

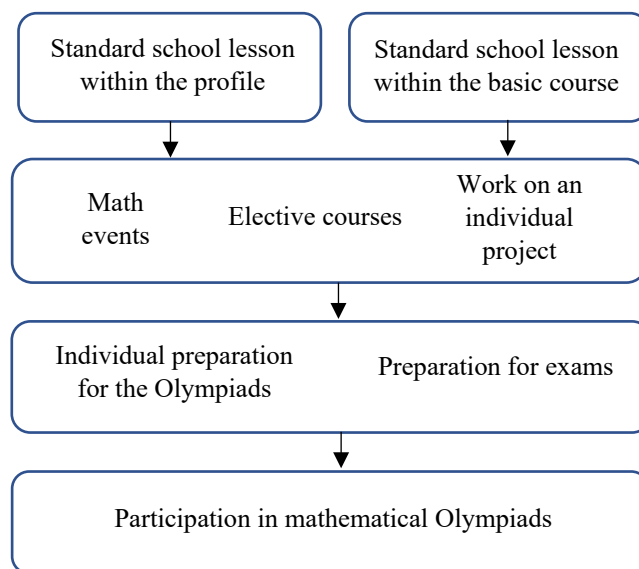


Figure 3: Relationship between different learning forms

## 7. Implementation of Model of Preparation of Gifted Children in Mathematics

The constructed model was used by us over a period of two years, working as mathematics teachers with gifted students at the Abulfazl Balami gymnasium for gifted children in the city of Vahdat, Republic of Tajikistan.

Let us briefly outline the strategy we employed.

First, during the lesson we observed how students solved partially algorithmic problems and identified which students were potentially capable of mathematics. These students were further involved in attending mathematical circles, where they tried their hand at school-level Olympiads.

If a student developed a sufficiently persistent interest in mathematics, and significant progress in the development of his abilities was evident, then he entered specialized classes. The persistent interest of students was seen in the process of students solving non-standard tasks [33]. The desire to solve such tasks, and, most importantly, the success in applying heuristic methods showed progress in the development of the abilities of the gifted students. Here they more purposefully, in parallel with the lessons of the profile course, continued their studies in the classroom of elective courses. For the most successful students, we developed a special learning plan that involved additional training activities and consultations in addition to elective classes.

In addition to the training of gifted schoolchildren immersed in subject mathematical activity, we made special efforts to ensure that all schoolchildren were constantly involved in joint communicative activities with classmates. In particular, in the classroom they could act as consultants on the subject, constantly participating in school competitions, concerts, sports events. This practice contributed to the development of the communicative abilities of schoolchildren, their emotional and volitional sphere, and reduced the risk of manifestation of possible difficulties in productive interaction with peers and adults around them.

Thanks to such an organization of the educational process, as our experience showed, non-standard abilities of schoolchildren developed and improved quite successfully, which was partially confirmed by the results of examinations and Olympiads.

In more detail, the methodological features of the implementation of this model are reflected in methodological materials we created and published in several articles.

Table 2 presents examples of some individual educational plans that we have built for gifted schoolchildren, as well as the first results of schoolchildren studying within the framework of the constructed model (Table 2).

Table 2: The Results of Work within the Framework of the Constructed Model

Student A - Grade 3 Works within the framework of the model for 1 year	Student B - Grade 4 Works within the framework of the model for 2 years
Forms of study: 1. standard school lesson within the framework of the basic course 2. elective courses 3. mathematical events Results: 1. shows interest in solving non-standard problems 2. Olympiads winner	Forms of study: 1. standard school lesson within the framework of the basic course 2. elective courses 3. individual preparation for the Olympiads 4. participation in mathematical Olympiads Results: 1. can solve problems of an increased level of complexity by various methods, choosing the most rational one 2. has high-level thinking according to Bloom's taxonomy, that is has high levels of cognitive ability 4. has a strong interest in learning mathematics 5. Olympiads winner
Student C- Grade 5 Works within the model for 2 years	Student D- Grade 6 Works within the model for 2 years
Forms of study: 1. standard school lesson within the framework of the basic course - 2. elective courses	Forms of study: 1. standard school lesson within the framework of the basic course - 2. elective courses 3. individual preparation for the Olympiads

3. individual preparation for the Olympiads 4. participation in mathematical Olympiads Results: 1. can solve problems of an increased level of complexity by various methods, choosing the most rational one 2. has a strong interest in the study of mathematics 3. goes to the final rounds of level Olympiads 4. is the winner and prize-winner of the final rounds of level Olympiads	Participation in mathematical Olympiads 4. preparation for the profile Results: 1. can solve problems of an increased level of complexity by various methods, choosing the most rational one 2. has high-level thinking according to Bloom's taxonomy, that is has high levels of cognitive ability 3. has a strong interest in advanced mathematics 4. goes to the final rounds of level Olympiads 5. is the winner and prize-winner of the final rounds of level Olympiads 6. claims for high scores in the profile
---	---

A pedagogical experiment to determine the possibilities of the proposed strategy for the training of gifted students, based on a rational combination of various formats of such training, was carried out over a two-year course of mathematical training of gifted schoolchildren in the Abulfazl Balami gymnasium for gifted children in the city of Vahdat, Republic of Tajikistan and a number of educational organizations in Dushanbe. In total, 41 students of the basic and senior levels of education were included in the experimental sample. The control groups in the study included two classes, in one of which (26 students) the dominant format for preparing gifted students was a series of elective courses devised by the authors. In the other the leading factor in training was individual consultations with specialists (15 students).

The data collection methodology included an analysis of the performance of current diagnostic work, including tasks of increased complexity, as well as taking into account the achievements of schoolchildren participating in the study in mathematical Olympiads of municipal, regional and republican status. In particular, the readiness of schoolchildren to solve non-standard tasks of increased difficulty using heuristic methods, their ability to find and compare different methods of performing the same task and to perform Olympiad tasks of a research nature were assessed. The evaluation of the results was carried out using the traditional five-point scale. The generalized result was considered as an average score for all diagnostic work. Students who received marks in the interval (2;3] were assigned to the first (low) level; in the interval (3;4] to the second (middle) level, and in the interval (4;5] - to the third (highest) level.

The received ordinal data at the pre- and post-implementation stages in the experimental and control groups were subjected to qualitative and quantitative analysis.

At the initial stage, the differences between the groups according to the selected levels proved to be unreliable. Statistical processing using the non-parametric fit method  $\chi^2$  - Pearson showed that the empirical values of the Pearson criterion when comparing the distributions of estimates in pairs in three samples proved to be lower than the corresponding critical values for given sample sizes. This fact indicates a relatively similar distribution of gifted schoolchildren by levels of success.

Experimental work within the framework of the ongoing study was carried out for two years. The experimental group studied according to the authors' model described above. For one control group the teacher used elective courses only. For the second control group the teacher used individual consultation only. The educational material in different formats in the control groups did not specifically correlate in any way either, in the program or in the procedural aspect.

As a result of experimental training, control measures were again carried out. A comparison of the dynamics of changes in success in the selected groups of students showed that over the same time interval in the experimental classes, about 40% of schoolchildren moved to a higher level of success, while the number and composition of students at each level in the second and third groups changed less significantly (Table 3).

Table 3: Results of Pedagogical Experiment

Level	After the experiment					
	Experimental group		Control group 1		Control group 2	
	Quantity	%	Quantity	%	Quantity	%
Initial	8	19	10	42,3	8	66,7
Medium	15	35,7	10	38,4	2	6,7
High	19	45,2	6	19,2	5	26,7

Statistical processing using the non-parametric fit method  $\chi^2$  - Pearson showed that the empirical values of Pearson's criterion in a pairwise comparison of the distributions of estimates in three samples turned out to be higher than the corresponding critical values for given sample sizes. When comparing the experimental sample and the first control sample, the following empirical value of the  $\chi^2$  --Pearson criterion was obtained  $\chi_e^2 = 8.2$ . The corresponding critical value at  $p \leq 0.05$  is significantly less than the empirical one (5.991). Similar results were obtained when comparing the experimental sample and the second control sample. Thus, after the use of the authors' model in the experimental group, students showed significant improvement in comparison with students in the two control group which used other approaches.

Due to the largely individual nature of work with gifted children, the generalized results of this diagnosis, in our opinion, cannot be considered an absolutely reliable indicator of the effectiveness of the study. Therefore, we also carried out an expert assessment of the study materials as an additional diagnostic technique. It was attended by 18 experienced mathematics teachers working in specialized mathematics classes. Their survey

showed that the vast majority of teachers confirmed the feasibility and prospects of the proposed methodological solutions.

## 8. Conclusion

The problem of training gifted students is now becoming particularly relevant. Purposeful provision of such training involves the development of a number of methodological solutions relating, in particular, to the rational correlation in the educational process of the relevant content, methods and teaching resources.

When considering this, we analyzed regulatory and policy documents, programs in mathematics of basic and elective courses for specialized mathematical classes, scientific and scientific-methodical works of leading domestic and foreign experts in the field of developmental psychology, didactics, theory and methodology of mathematical education, as well as existing textbooks, teaching aids, methodological recommendations and software for educational purposes. In addition, a longitudinal observation was made of mathematics courses for gifted schoolchildren in a number of educational institutions and a survey was conducted of mathematics teachers working in specialized classes. The survey revealed the difficulties that arise when studying in these classes. As a result of this work, it was discovered that the majority of the teachers surveyed indicated, for the most part, that there is insufficient interaction of various formats of mathematical training of gifted students, which does not ensure its integrity.

The following relatively new results were obtained.

1. The authors' model of teaching gifted schoolchildren in mathematics has been developed and theoretically substantiated. The model includes components that reflect the content, forms and methods of working with gifted children in the framework of the main and optional courses.
2. The main methods of working with gifted children in mathematics classes which ensure their active search and research and creative activity are disclosed. The "mechanism of their interaction" in the educational process is also disclosed.
3. A holistic strategy for the work of a mathematics teacher with gifted children has been determined, encompassing classes within the framework of basic and elective mathematical courses, work on individual projects, participation in mathematical holidays and preparation for mathematical Olympiads. All of these formats are integrated into the individual educational routes of each of the students, ensuring the quality of their mathematical preparation and a high level of intellectual development.
4. Various options for constructing individual educational routes were identified and tested in the implementation of the developed strategy, depending on the stage of teaching mathematics. The implementation of these options formed the basis for the development of methodological materials and recommendations for the preparation of gifted students, taking

into account the continuity of basic and optional mathematical courses.

A pedagogical experiment to determine the possibilities of the proposed strategy for the training of gifted students based on a rational combination of various formats of such training, was carried out in the course of a two-year subject mathematical training of gifted schoolchildren in the Abulfazl Balami gymnasium for gifted children in the city of Vahdat (Republic of Tajikistan) and a number of educational organizations in Dushanbe. During the experiment, the educational and developmental capabilities of three models of mathematical training of gifted children were compared: one experimental group and two control groups, in which the relationship of related formats of teaching mathematics was not specifically taken into account. As a result of experimental training, a pairwise comparison of the degree of change in success in the selected groups of students showed that for the same time interval in the experimental classes, 40 % of the students moved to a higher level of success, while the number and composition of students at each level in the control groups changed less significantly.

In general, it can be concluded that empirical learning based on the methodological determinants and recommendations outlined above proved to be feasible and quite effective within the framework of the current regulatory formats. This is evidenced, in particular, by a fairly large number of schoolchildren participating in Olympiads of various levels, and high scores in examinations. This result was also confirmed by the results of an expert evaluation of the proposed methodological solutions by mathematics teachers working in specialized mathematical classes.

As a further development of our work, we are considering the development of an adaptive pedagogical technology for working with gifted students, which will contribute to the development of their mathematical abilities, as well as provide for effective learning to solve Olympiad problems. For teachers, this technology will help build an individual learning plan for each gifted student, naturally updating his cognitive activity.

## References

- [1] A. G. Balm, "The Effects of Discovery Learning on Students' Success and Inquiry Learning Skills," *Eurasian Journal of Educational Research*, Issue 35, 1-20 Spring 2009, 2009.
- [2] L. Alfieri, P.L. Brooks, N.J. Aldrich, H. R. Tenenbaum, "Does discovery-based instruction enhance learning?" *Journal of Educational Psychology*, **103**(1), 1–18, 2011, doi.org/10.1037/a0021017.
- [3] J. Gallagher, *Teaching the gifted child* (2nd ed.). Boston: Allyn and Bacon, 1975.
- [4] S. Assouline, A. Lupkowski-Shoplik, *Developing Math Talent*. Texas: Prufrock Press, 2011.
- [5] F. Barron, *Creativity and the gifted*. In *New directions for gifted education*. Report on bicentennial mid-year Leadership Training Institute. Los Angeles: National/State Leadership Training Institute on the Gifted and Talented, 1976, doi.org/10.1177/001698628002400306
- [6] C.P. Benbow, L. L. Minor, "Cognitive profiles of verbally and mathematically precocious students: Implications for identification of the gifted," *Gifted Child Quarterly*, **34**(1), 21–26, 1990, doi.org/10.1177/001698629003400105.
- [7] D. Bogoyavlenskaya, V. Shadrikov *Working concept of giftedness*. (2nd ed) M., Progress, 2003.
- [8] G. Davis, S. Rimm, *Education of the gifted and talented* (5th ed.). Boston, MA: Allyn & Bacon, 2004.
- [9] F. Gagné, "Transforming gifts into talents: The DMGT as a developmental theory," In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (3rd ed., 60–74). Boston, MA: Allyn & Bacon, 2003.
- [10] L. Vygotsky, *Imagination and creativity in childhood*. St. Petersburg, SOYUZ, 1997.
- [11] M. Hoeflinger, "Developing mathematically promising students," *Roeper Review*, **20**(4), 244–247, 1998, doi.org/10.1080/02783199809553900.
- [12] E. P. Torrance, *Guiding creative talent*. Englewood Cliffs, N. J.: Prentice-Hall, 1962.
- [13] G. A. Goldin, "Mathematical creativity and giftedness: perspectives in response," *ZDM* **49**, 147–157, 2017 doi.org/10.1007/s11858-017-0837-9.
- [14] F. Barron, *Creative Person and Creative Process*, Holt, Rinehart & Winston, New York, 1969.
- [15] J. Gilford, *Three sides of the intellect*. Psychology of thinking. M., Progress, 1965.
- [16] C. W. Taylor, "Cultivating simultaneous student growth in both multiple creative talents and knowledge," In J. S. Renzulli (Ed.), *Systems and models for developing programs for the gifted and talented*, 306–351, 1986.
- [17] E. P. Torrance, "Growing up creatively gifted: A22-year longitudinal study," *Creative Child and Adult Quarterly*, **5**(3), 148–158, 1980.
- [18] E. Torrance, J. Khatena, "Originality of imagery in identifying creative talent in music," *Gifted Child Quarterly*, **13**, 3-8, 1969, doi.org/10.1177/001698626901300101.
- [19] G. Lombroso, *The Man of Genius*. London: W. Scott, 1891.
- [20] J. Guilford, *The nature of intelligence*. New York: McGraw-Hill, 1967.
- [21] S. Rubinshtein, *Fundamentals of general psychology*. M, Uchpedgiz. 1940.
- [22] A. Karp, "Knowledge as a manifestation of talent: Creating opportunities for the gifted," In B. Sriraman (Ed.), *Creativity, giftedness, and talent development in mathematics* (209–224). Charlotte, NC: Information Age Publishing, 2008, <https://www.diva-portal.org/smash/get/diva2:1390686/FULLTEXT01.pdf>.
- [23] T. Hirano, "Achieving mathematical excellence in Japan: Results and implications," *Journal of Educational Research*, **25**(6), 545-551, 1996, doi.org/10.1016/S0883-0355(97)86731-6.
- [24] K. Heller, A. Lengfelder, *German Olympiad study on math, physics and chemistry*. Paper presented at the American Educational Research Association, New Orleans, 2000, doi.org/10.1080/02783193.2011.530202.
- [25] C. E. Hmelo-Silver, H. S. Barrows, *Goals and Strategies of a Problem-based Learning Facilitator*. *Interdisciplinary Journal of Problem-Based Learning*, **1**(1), 2006, doi.org/10.7771/1541-5015.1004.
- [26] R. Campbell, H. Walberg, "Olympiad Studies: Competitions Provide Alternatives to Developing Talents That Serve National Interests," *Roeper Review*, **33**(1), 8-17, 2011, doi.org/10.1080/13803610701785949
- [27] Y. Terada, *Boosting student engagement through project-based learning*. Edutopia, 2018.
- [28] K. Tirri, "Finland Olympiad Studies: What factors contribute to the development of academic talent in Finland," *Journal of NACE.*, **5**(2), 56-66, 2001.
- [29] J. Boaler, "Promoting 'relational equity' and high mathematics achievement through an innovative mixed-ability approach," *Br. Educ. Res. J.* **34**, 167–194, 2008 doi.org/10.1080/01411920701532145.
- [30] D. P. Wolf, *The art of questioning*. Academic Connections, 1987.
- [31] A. Benson, D. Blackman, "Can research methods ever be interesting?," *Active Learning in Higher Education* **4**(1), 39-55, 2003, doi.org/10.1177/1469787403004001859.
- [32] C. H. Chen, Y. C. Yong, "Revisiting the effects of project-based learning on students' academic achievement: A meta-analysis investigating moderators," *Educational Research Review*, **26**, 71–81, 2019, <https://www.learnlib.org/p/207141/>.
- [33] S. Cho, H. Lee, "Korean gifted girls and boys: What influenced them to be Olympians and non Olympians," *Journal of Research in Education*, **12**(1), 106–111, 2002.
- [34] D. Kokotsaki, V. Menzies, A. Wiggins, "Project-based learning: a review of the literature," *Improving schools.*, **19**(3). pp. 267-277, 2016, doi.org/10.1177/1365480216659733.
- [35] A. Mettas, C. Constantinou, "The Technology Fair: a project-based learning approach for enhancing problem solving skills and interest in design and technology education," *International Journal of Technology and Design Education*, **18**, 79-100, 2007, doi.org/10.1007/s10798-006-9011-3.
- [36] J. Brunstein, *Achievement motivation* (H. Heckhausen, Ed.). In J. Heckhausen & H. Heckhausen (Eds.), *Motivation and action* (137–183). Cambridge University Press, 2008,

doi.org/10.1017/CBO9780511499821.007

- [37] A. Conley, "Patterns of motivation beliefs: combining achievement goal and expectancy-value perspectives," *Educ. Psychol.* **104**, 32–47, 2012, doi.org/10.1037/a0026042
- [38] C. S. Dweck, *Self-theories: Their role in motivation, personality and development*. Philadelphia, Psychology Press, 1999.
- [39] S. K.W. Chu, S. K. Tse, K. Chow, "Using collaborative teaching and inquiry project-based learning to help primary school students develop information literacy and information skills," *Library & Information Science Research*, 33, 132-143, 2011, doi.org/10.1016/J.LISR.2010.07.017.