IIUM ENGINEERING JOURNAL

Volume 26 Number 3

September 2025

IIUM Press

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

ISSN: 1511-788X E-ISSN: 2289-7860

http://journals.iium.edu.my/ejournal

IIUM ENGINEERING JOURNAL

Volume 26, Issue 3, September 2025

https:/doi.org/10.31436/iiumej.v26i3

Copyright Notice

Consent to publish: The Author(s) agree to publish their articles with IIUM Press.

Declaration: The Author(s) declare that the article has not been published before in any form, that it is not concurrently submitted to another publication, and that it does not infringe on anyone's copyright. The Author(s) holds the IIUM Press and Editors of the journal harmless against all copyright claims.

Transfer of copyright: The Author(s) hereby agree to transfer the article's copyright to IIUM Press, which shall have the non-exclusive and unlimited right to publish the article in any form, including in electronic media. For articles with more than one author, the corresponding author confirms that he/she is authorized by his/her co-author(s) to grant this copyright transfer.

The IIUM Engineering Journal follows the open access policy.

All articles published for open access will be immediately and permanently free for everyone to read, download, copy, and distribute for noncommercial purposes.

IIUM Engineering Journal at https://journals.iium.edu.my/ejournal is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

IIUM ENGINEERING JOURNAL EDITORS

CHIEF EDITOR
Ahmad Faris Ismail, IIUM, Malaysia

TECHNICAL EDITOR
Khairul Azami Sidek, IIUM, Malaysia

EXECUTIVE EDITOR
Teddy Surya Gunawan, IIUM, Malaysia

ASSOCIATE EDITOR
Nor Farahidah Za'bah, IIUM, Malaysia

COPY EDITOR
Aliza 'Aini Md Ralib, IIUM, Malaysia
Suriza Ahmad Zabidi, IIUM, Malaysia

MALAY TRANSLATOR
Nurul Arfah Che Mustapha, IIUM, Malaysia

EDITORIAL BOARD MEMBERS

Abdullah Al-Mamun, IIUM, Malaysia
Abdumalik Rakhimov, IIUM, Malaysia
Aishah Najiah Dahnel, IIUM, Malaysia
Alya Naili Rozhan, IIUM, Malaysia
Ani Liza Asnawi, IIUM, Malaysia
Hanafi Ani, IIUM, Malaysia
Hanafy Omar, Saudi Arabia
Huda Adibah Mohd. Ramli, IIUM, Malaysia
Konstantin Khanin, University of Toronto, Canada
Ma'an Al-Khatib, IIUM, Malaysia
Meftah Hrairi, IIUM, Malaysia
Mohamed B. Trabia, United States
Mohammad S. Alam, Texas A&M University-Kingsville, United States
Mohd Sultan Ibrahim Shaik Dawood, IIUM, Malaysia
Mustafizur Rahman, National University Singapore, Singapore
Noorasikin Samat, IIUM, Malaysia
Nor Fadhillah Mohamed Azmin, IIUM, Malaysia
Norsinnira Zainul Azlan, IIUM, Malaysia
Ossama Abdulkhalik, Michigan Technological University, United States
Waqar Asrar, IIUM, Malaysia

AIMS & SCOPE OF IIUMENGINEERING JOURNAL

The **IIUM Engineering Journal**, published biannually (January and July), is a carefully refereed international publication of the International Islamic University Malaysia (IIUM). Contributions of high technical merit within the span of engineering disciplines covering the main areas of engineering: Electrical and Computer Engineering; Mechanical and Manufacturing Engineering; Automation and Mechatronics Engineering; Material and Chemical Engineering; Environmental and Civil Engineering; Biotechnology and Bioengineering; Engineering Mathematics and Physics; and Computer Science and Information Technology are considered for publication in this journal. Contributions from other areas of engineering and applied science are also welcome. The IIUM Engineering Journal publishes contributions under *Regular papers and Invited review papers*. It also welcomes contributions that address solutions to the specific challenges of the developing world, as well as science and technology issues, from an Islamic and multidisciplinary perspective.

REFEREES' NETWORK

All papers submitted to the IIUM Engineering Journal will be subjected to a rigorous reviewing process through a worldwide network of specialized and competent referees. Each accepted paper should have at least two positive referees' assessments.

SUBMISSION OF A MANUSCRIPT

A manuscript should be submitted online to the IIUM-Engineering Journal website at https://journals.iium.edu.my/ejournal. The journal website can also be used for further correspondence on the paper's status.

INTERNATIONAL ADVISORY COMMITTEE

A. Anwar, United States
Abdul Latif Bin Ahmad, Malaysia
Farzad Ismail, USM, Pulau Pinang, Malaysia
Hanafy Omar, Saudi Arabia
Hany Ammar, United States
Idris Mohammed Bugaje, Nigeria
K.B. Ramachandran, India
Kunzu Abdella, Canada
Luis Le Moyne, ISAT, University of Burgundy, France
M Mujtaba, United Kingdom
Mohamed AI-Rubei, Ireland
Mohamed B Trabia, United States
Syed Kamrul Islam, United States
Tibor Czigany, Budapest University of Technology and Economics, Hungary
Yiu-Wing Mai, The University of Sydney, Australia.

Published by:

IIUM Press

IIUM Press,
International Islamic University Malaysia
Jalan Gombak, 53100 Kuala Lumpur, Malaysia
Phone (+603) 6421-5014, Fax: (+603) 6421-6298

Whilst every effort is made by the publisher and editorial board to see that no inaccurate or misleading data, opinion or statement appears in this Journal, they wish to make it clear that the data and opinions appearing in the articles and advertisement herein are the responsibility of the contributor or advertiser concerned. Accordingly, the publisher and the editorial committee accept no liability whatsoever for the consequence of any such inaccurate or misleading data, opinion, or statement.

ISSN 1511 - 788X

9 771511 788008

IIUM Engineering Journal
ISSN: 1511-788X E-ISSN: 2289-7860

https://doi.org/10.31436/iiumej.v26i3.3918

EDITORIAL MESSAGE

The IIUM Engineering Journal Vol. 26 No. 3 continues its commitment to advancing impactful research that addresses contemporary engineering challenges through innovation, sustainability, and interdisciplinary integration. This issue features 28 high-quality papers across a broad spectrum of disciplines: Civil and Environmental Engineering, Electrical, Computer and Communications Engineering, Engineering Mathematics and Applied Science, Materials and Manufacturing Engineering, and Mechatronics and Automation Engineering.

The Civil and Environmental Engineering section reflects a growing focus on safety, durability, and sustainable infrastructure. Papers in this category explore accident probability among elderly motorcyclists in Indonesia, corrosion assessment in fly ash–silica fume concrete via non-destructive testing, eco-processed pozzolan characterization, strength forecasting using machine learning and response surface methodology, clayey sand improvement using brick dust, and sediment scour simulations in submerged weirs. Collectively, these works contribute to improved construction practices and enhanced environmental resilience.

The Electrical, Computer, and Communications Engineering section comprises the largest set of contributions, showcasing cutting-edge AI and computational advancements across multiple domains. Topics include real-time deep learning for UAV object detection, electronic nose integration with LeNet and regularization, blockchain benchmarking on heterogeneous IoT hardware, and health monitoring of lead-acid batteries in off-grid solar systems. Other notable studies investigate fake news detection using multimodal embeddings, hybrid deep learning for data center temperature control, sensor-based fingerspelling recognition, and spatiotemporal models for earthquake prediction. Additionally, papers on chipless RFID resonators, LoRa-driven UAV return prediction, saliency-based segmentation of skin cancer, and sentiment analysis on TVET from social media highlight AI's broad impact. Further innovations are presented in malware detection within IoMT environments, trade-space analysis of satellite anomalies, parametric cost modeling, and power generation optimization using bio-inspired algorithms. These papers demonstrate the transformative role of AI, IoT, and intelligent systems in reshaping engineering practice.

In Engineering Mathematics and Applied Science, a significant theoretical contribution explores a novel seven-dimensional chaotic system, enriching our understanding of nonlinear dynamics and complex systems. The Materials and Manufacturing Engineering section highlights progress in safety engineering and the development of advanced composites. Featured studies include the enhanced design of portable oil spill skimmers using computational fluid dynamics, void formation analysis in BFS/$CaCO_3$ diffusion couples, and the evaluation of machining performance in hybrid fiber-reinforced polymer composites. These contributions reflect ongoing efforts to optimize material performance and environmental protection.

In Mechatronics and Automation Engineering, contributions emphasize intelligent control and autonomous systems. Topics include the dynamic analysis of aerial work platforms using hybrid CAD for satellite testing, as well as autonomous navigation based on frontier-based detection integrated with social force modeling. These efforts highlight the integration of robotics, control systems, and intelligent automation to improve operational reliability and mobility in complex environments.

This issue exemplifies the journal's mission to bridge theory and real-world application across diverse engineering domains, promoting secure, sustainable, and intelligent systems. The editorial team expresses its sincere gratitude to all authors, reviewers, and section editors for their valuable contributions, commitment to excellence, and support in upholding the scholarly integrity of this publication. We hope these contributions inspire future innovations, foster stronger academia–industry collaboration, and accelerate global progress towards sustainable, inclusive, and technology-driven development.

Prof. Ir. Ts. Dr. Teddy Surya Gunawan

Executive Editor

IIUM Engineering Journal

IIUM ENGINEERING JOURNAL

Volume 26, Issue 3, September 2025
https:/doi.org/10.31436/iiumej.v26i3

Table of Contents

Copyright i
Editorial Board ii
Editorial Message iv

Civil and Environmental Engineering

3474: Probability of Single-Vehicle Accidents Among Elderly Motorcyclists in Indonesia 1
Pada Lumba, Ahmad Fathoni, Arifal Hidayat, Anton Ariyanto, Harriad Akbar Syarif

3606: Corrosion Evaluation on Fly Ash and Silica Fume Concrete Using NDT Methods..... 16
Ahmad Zaki, Ferdi Ardian Syah, Sri Atmaja P. Rosyidi, Kharisma Wira Nindhita, Zainah Ibrahim

3631: Characterisation and Strength Activity Index of Eco-Processed Pozzolan 41
Abraham Henry Bernard, Ahmad Nurfaidhi Rizalman

3730: Machine Learning and RSM for Strength Forecasting in Sustainable SCGC 53
Sameh Fuqaha, Ahmad Zaki, Guntur Nugroho

3821: Initial Investigation on Improving the Physicomechanical Properties of Clayey Sand Using Clay Brick Dust 86
Umme Saima Nur, Norhidayu binti Kasim, Wan Nur Aifa Wan Azahar

3825: Numerical Investigation of Sediment Scour Downstream of Adjustable and Fixed Submerged Weir Using FLOW-3D 102
Siti Nurain Che Mohd Azmi, Saerahany Legori Ibrahim, Izihan Ibrahim, Dani Irwan Masbah

Electrical, Computer, and Communications Engineering

3433: Designing and Implementing Real-Time Deep Learning Object Detection in Unmanned Aerial Vehicles 123
Nurul Fahmi Arief Hakim, Alyani Ismail, Iwan Kustiawan, Deasy Rosanti, Taqiyuddin Yazid Zaidan

3471: An Integration of LeNet with Regularization Techniques for Electronic Nose in Air Contaminant Classification 138
Muhammad Arif Syahmi Md Dzahir, Kim Seng Chia

3610: Performance Benchmarking of Hyperledger Fabric on Heterogeneous Hardware for IoT Applications 156
Muhammad Muaz Zulkarnain, Nabilah Ramli, Anis Nurashikin Nordin

3612: Challenges and Strategies for Implementation of Lead-Acid Battery Health Monitoring and Prediction in Off-Grid Solar Panel Systems..... 171
Yoga Prastiya Wibawa, I Putu Ananta Yogiswara, Bondan Suwandi, Andi Kurnianto, Budi Sulistya, Abhimata Ar-Rasyiid, Nashrullah Taufik, Arief Rufiyanto, I Made Astawa

3615: A Detector for Textual-Visual Fake News Using Text Summarization and Contrastive Language-Image Pretraining Embedding Model 185
Iman Qays Abduljaleel, Israa H. Ali

3629: Development of a Predictive Real-Time Temperature Management Framework for Data Centers Using a Hybrid Deep Learning Model 200
Patrick Olabisi, Ahmed Mayowa

3633: Modelling of Flex Sensor Resistance Value for Fingerspelling Recognition Prototype Using Sensor-Based Signal 220
Muhammad Hazmirul Haqim Hazri, Zarina Mohd Noh, Norhidayah Mohamad Yatim, Nik Mohd Zarifie Hashim

3634: Proposed ConvBiLSTM-Net Model for Enhancing Earthquake Prediction Performance Using Spatiotemporal Features 238
Ari Fadli, Tri Kuntoro Priambodo, Agfianto Eko Putra, Wiwit Suryanto

3646: Miniaturized L-Shaped and U-Shaped Resonator-Based 8-Bit and 12-Bit Chipless RFID Tag 260
Sanaa Salama, Omar Tamimi, Munerra Altayeb, Ahlam Abuzahew

3652: LoRa-Driven Algorithms for Accurate Return-to-Home (RTH) Prediction in Unmanned Aerial Vehicles 270
Mohamad Hazwan Mohd Ghazali, Han Heng, Izzati Saleh, Wan Rahiman

3680: Boundary-Aware Saliency-Based Level Set with Momentum Contrast Metaformer for Skin Cancer Segmentation and Classification 280
Anjani Gupta, Snehlata Sheoran, Surabhi Rastogi, Anshula Gupta, Kajal Saluja, Sabnam Kumari

3739: Identifying Technical and Vocational Education and Training (TVET) Sentiment from Social Media Using a Machine Learning Approach 295
Nur Hafazah Sharin, Mira Kartiwi

3746: A Novel Hybrid Model for High-Accuracy Malware Detection in the Internet of Medical Things (IoMT) Environment 304
Amarudin Daulay, Kalamullah Ramli, Dodi Sudiana, Ruki Harwahyu, Taufik Hidayat, Ernia Susana, Nurwan Reza Fachrurrozi

3788: Reliability Trade-Space Exploration Modelling for Satellite Anomalies Using Exponential Distribution 320
Muhammad Imran Haris Ahmad Mahmood, Nadirah Abdul Rahim, Abdul Mutholib

3794: Trade-Space Exploration Comparison of Parametric Cost Models for Satellite Anomalies with RMSE and RRMSE 331
Muhammad Aryss Zafran Shahrul Ariffin, Nadirah Abdul Rahim, Abdul Mutholib

3800: Optimizing Generation Cost and Reducing Gas Emissions in Power Generation Using the Artificial Bee Rabbit Optimization Algorithm 343
Wei Wen Lee, Mohd Ruzaini bin Hashim, Chin Kim Gan

Engineering Mathematics and Applied Science

3590: Construction and Analysis of a New Seven-Dimensional Chaotic System with Complex Dynamical Behaviors 360
Sura Hasballah Shnawa, Sadiq A. Mehdi, Emad I. Abdul Kareem

Materials and Manufacturing Engineering

3253: Safety Enhancement of Portable Oil Spill Skimmer (POSS) via Computational Fluid Dynamics for Liquid Sloshing Analysis 371
Rprakash Ramanathan, Lokman Abdullah, Muhammad Hafidz Fazli Md Fauadi, Muhammad Syafiq Syed Mohamed, Ruzaidi Zamri, Rahimah Abdul Hamid, Zamberi Jamaludin, and Mohd Shahrieel Mohd Aras

3603: Formation of Void in BFS/$CaCO_3$ Diffusion Couple 386
Ahmad Abdul Mun'im Ismail, Alya Naili Rozhan, Muhammad Rafiq Haikal Rosdin, Hadi Purwanto, Abd Malek Abdul Hamid, M. H. Ani

3671: Machining Performance with Respect to Cutting Forces, Vibrations, and Surface Quality in Drilling of Hybrid Abaca and Glass Fibers Reinforced Polymer Composite . 397
Mohd. Iqbal, Teuku Firsa, Nazaruddin, Indera Sakti Nasution, Muhammad Rizal, Jihadul Qahhar, Irfan Hilmy

Mechatronics and Automation Engineering

3497: Dynamic Analysis on an Aerial Work Platform Using a Hybrid CAD Approach for Satellite Testing Applications 412
Nor Mohd Haziq Norsahperi, Salmiah Ahmad, Siti Fauziah Toha, Mohd Azri Abd Mutalib, Mohamad Ridzwan Ishak, Dong Eui Chang
3554: Frontier-Based Detection and Social Force Model for Autonomous Environment Mapping and Navigation..... 430
Asyam Irsyad, Bima Sena Bayu Dewantara, Setiawardhana

Blank page.

https://doi.org/10.31436/iiumej.v26i3.3474

PROBABILITY OF SINGLE-VEHICLE ACCIDENTS AMONG ELDERLY MOTORCYCLISTS IN INDONESIA

PADA LUMBA[1*], AHMAD FATHONI[2], ARIFAL HIDAYAT[1], ANTON ARIYANTO[1],HARRIAD AKBAR SYARIF[1]

[1]*Civil EngineeringDepartment, Faculty of Engineering,Universitas Pasir Pengaraian, Pasir Pengaraian, 28457, Riau Province, Indonesia*

[2]*Mechanical Engineering Department, Faculty of Engineering,Universitas Pasir Pengaraian, Pasir Pengaraian, 28457, Riau Province, Indonesia*

**Corresponding author: padalumba@gmail.com*

(Received: 18 October 2024; Accepted: 21 March 2025; Published online: 9 September 2025)

ABSTRACT: The number of accidents involving elderly motorcyclists is relatively high compared to other age groups. This is due to various limitations commonly experienced by older riders. This study aims to determine the probability of single-vehicle accidents among elderly motorcyclists in relation to human, road, and environmental factors. A total of 564 respondents participated in the study conducted in Riau Province, Indonesia. Data were collected through interviews with elderly motorcyclists who had previously experienced accidents. The data were analyzed using a Bayesian Network model with GeNIe 2.0 software. The results showed that the probability of single-vehicle accidents among elderly motorcyclists is 59%. Model validation indicated a Mean Absolute Deviation (MAD) of 23%. Male elderly motorcyclists have a 60% likelihood of experiencing single-vehicle accidents (Scenario 1), while female elderly motorcyclists have a 59% likelihood (Scenario 2). Those who ride while fatigued have a 65% probability of a single-vehicle accident (Scenario 3), compared to 57% for those who are not fatigued (Scenario 4). Riding in rainy conditions increases the probability to 70% (Scenario 5), whereas riding in dry conditions reduces it to 55% (Scenario 6). Elderly motorcyclists riding on potholed roads have a 64% chance of accidents (Scenario 7), compared to 57% on roads without potholes (Scenario 8). These findings indicate that elderly riders are highly vulnerable to single-vehicle accidents. Among human factors, fatigue is the most significant variable influencing accident probability. Regarding environmental factors, driving in the rain plays a key role, while riding on potholed roads is the primary influence for road factors. This study highlights the dominant factors contributing to single-vehicle accidents among elderly motorcyclists by integrating human, road, and environmental considerations.

ABSTRAK: Bilangan kemalangan dalam kalangan penunggang motosikal warga emas agak tinggi berbanding kumpulan umur yang lain.Hal ini disebabkan oleh beberapa batasan yang dialami oleh penunggang motosikal warga emas.Tujuan kajian ini adalah untuk menentukan kebarangkalian kemalangan bujang dalam kalangan penunggang motosikal warga emas berkaitan faktor manusia, jalan raya dan persekitaran.Jumlah sampel ialah 563 responden dan lokasi kajian adalah di Wilayah Riau, Indonesia. Pengumpulan data yang pernah dialami oleh penunggang motosikal dengan menggunakan analisis elderata. Rangkaian Bayesian dengan Perisian GeNie 2.0. Keputusan menunjukkan kebarangkalian kemalangan bujang dalam kalangan penunggang motosikal warga emas ialah 59%.Hasil pengesahan model menunjukkan nilai MAD sebanyak 23%. Penunggang motosikal warga emas lelaki berkemungkinan mengalami kemalangan bujang sebanyak 60% (senario 1), penunggang motosikal warga emas wanita sebanyak 59% (senario 2). Penunggang motosikal warga emas yang memandu semasa keletihan berkemungkinan mengalami satu kemalangan sebanyak

65% (senario 3), dalam keadaan tidak letih 57% (senario 4 yang memandu dalam keadaan hujan yang berkemungkinan besar dalam keadaan hujan). 70% (senario 5), dalam keadaan tidak hujan 55% (senario 6). Penunggang motosikal warga emas yang memandu di jalan berlubang berkemungkinan akan mengalami satu kemalangan sebanyak 64% (senario 7), di jalan tanpa jalan berlubang 57% (senario 8). Bermakna pemandu warga emas sangat terdedah untuk mengalami kemalangan bujang. Pembolehubah yang paling mempengaruhi kemungkinan pemandu warga emas mengalami kemalangan bujang ialah keletihan dari segi faktor manusia. Bagi faktor persekitaran, pembolehubah yang mempengaruhi pemandu warga emas yang mengalami kemalangan bujang ialah memandu dalam hujan. Bagi faktor jalan raya, pembolehubah yang mempengaruhi kemalangan tunggal ialah memandu di jalan berlubang. Dapatan kajian ini mendapat faktor dominan yang menyebabkan kemalangan bujang pada pemandu warga emas dengan mengambil kira faktor manusia, jalan raya dan persekitaran.

KEYWORDS:*Accident, Bayesian, Elderly, Motorcyclists, Rider.*

1. INTRODUCTION

One of the characteristics of drivers that influences driving behavior is age [1-3]. As drivers age, their driving abilities and skills tend to decline, especially among older individuals. Young drivers generally have better physical and cognitive functions than adults and elderly drivers. However, with increasing age, there is a decline in both physical and cognitive functions. Elderly drivers may experience several limitations, including impaired vision, reduced hearing, difficulty concentrating, and trouble remembering routes and traffic signs. These limitations can significantly impair an elderly driver's ability to respond quickly to sudden warnings or dangers. The decline in driving ability among older drivers is caused by several factors, including reduced visual clarity, decreased concentration, lower stamina or increased fatigue [4], slower response times [5], difficulty understanding vehicle technology [6–9], and medical conditions that may affect driving performance [1,10,11].

Poor vision can make it difficult for elderly drivers to understand traffic in front of or around them. In such conditions, they may be unable to make quick and accurate decisions, such as whether to brake, swerve, or continue driving, significantly increasing the risk of an accident. A lack of concentration while driving also heightens this risk, as elderly drivers may take longer to respond to sudden warnings or dangers. Additionally, elderly drivers generally have lower stamina than younger drivers, particularly teenagers. As a result, they tire more quickly, leading to a decline in driving performance and an increased likelihood of accidents. Another challenge elderly drivers face is difficulty understanding modern vehicle technology, especially when operating newer models. Their inability to use current technology effectively can prevent them from operating the vehicle optimally, raising the risk of accidents. Limited vision while driving can prevent drivers from correctly understanding their surroundings, leading to misidentification of potential hazards and inappropriate responses. These conditions can increase the risk of accidents. Limited visibility while driving is one of the causes of accidents because drivers may not fully understand the situation in front of or around them, leading to misidentification of problems and inappropriate responses.

Reduced concentration while driving can lower a driver's level of alertness, making it harder to detect nearby dangers. This often leads to inaccurate decision-making, which can, in turn, cause accidents. Older drivers are also more prone to fatigue, which slows their ability to respond to sudden hazards. This condition increases the likelihood of accidents [12] and can also worsen their severity [13]. Driver fatigue is influenced by several factors, including a

heavy workload before driving, lack of sleep the night before [14], extended driving durations [15], and the time of day driving occurs [16]. Sleep deprivation contributes to fatigue [14], which negatively impacts driver performance [17, 18], ultimately increasing the risk of accidents. Additionally, when drivers are not fully concentrated, their level of alertness decreases, making it more difficult for them to detect potential dangers around them. As a result, their decisions may be inaccurate, increasing the risk of accidents. Furthermore, older drivers tend to experience fatigue more quickly, slowing their response to hazards and raising the risk of accidents [12]. Fatigue can also contribute to the severity of accidents [13]. Driver fatigue may be caused by several factors, including a heavy workload before driving, a lack of sleep the night before [14], extended periods of driving [15], and the time of day the driving occurs [16]. In particular, a lack of sleep can lead to fatigue [14], which negatively affects driver performance [17, 18] and increases the likelihood of accidents.

All of the above factors are important when driving, especially in preventing the risk of accidents. Another significant factor influencing the risk of accidents among elderly drivers is the driver's gender. Male drivers over 80 and female drivers over 70 are more likely to be involved in single-vehicle accidents [19]. Research also indicates that elderly male drivers have a higher probability of fatal single-vehicle accidents [20]. In 2020, the number of accident victims by age group was as follows: 26,906 for ages 10-19, 29,281 for ages 20-29, 18,553 for ages 30-39, 17,980 for ages 40-49, and 31,740 for those aged 50 and above. The number of accident victims aged 50 and older exceeds that of younger drivers. Furthermore, drivers who complete courses before the age of 75 have a lower risk of accidents [21]. Other studies also suggest that elderly drivers benefit more from understanding their physical and cognitive limitations than merely memorizing traffic laws [22]. This highlights the importance of self-awareness in older drivers over knowing highway regulations.

In addition to human factors, road conditions also impact the driving abilities of elderly drivers, including driving on straight roads, curves [8], intersections [23], and roads with slight roadside variation [24]. Driving on long, straight roads often requires minimal actions, such as braking or accelerating, while drivers maintain a constant speed. This can lead to increased fatigue and drowsiness, reducing driver alertness. Similarly, roads with monotonous surroundings, such as stretches bordered only by rice fields, can cause boredom and decreased focus. Elderly drivers tend to respond more slowly at intersections, which can reduce traffic flow capacity. While long straight roads may induce sleepiness and increase accident risk, drivers usually stay more alert on curves because they must carefully monitor road and traffic conditions [14, 24]. Poorly visible potholes, especially at night, can cause drivers to lose vehicle control and lead to single-vehicle accidents. Rain also limits visibility, making driving especially dangerous for elderly drivers.

In addition to the factors mentioned above, vehicle-related factors also influence the driving behavior of elderly drivers. These include incompatibility between the driver's body posture and the vehicle, limited understanding of in-vehicle technology, weather conditions, and poor or inadequate vehicle lighting. A car that does not suit the driver's posture can cause discomfort while driving, affecting the driver's behavior, especially during maneuvers. Furthermore, limited ability to understand and operate modern vehicle technology can make it difficult for elderly drivers to control the vehicle, potentially endangering their safety.

Statistical data from 2020 in Indonesia indicate that the accident rate is high among young drivers. As drivers age, the accident rate decreases significantly; however, it rises again in older age groups [25–27]. In that year, the number of traffic accident victims aged 10–19 was 26,906, those aged 20–29 reached 29,281, followed by 18,553 victims aged 30–39, and 17,980 aged 40–49. Surprisingly, the number of victims aged 50 and above was the highest at 31,740. This

https://doi.org/10.31436/iiumej.v26i3.3474

indicates that elderly drivers are involved in more accidents than both adolescents and middle-aged adults. In Indonesia, motorcycles are popular not only among young drivers but also among middle-aged and elderly riders. A survey conducted in 2024 among motorcyclists revealed that those who had been involved in accidents and showed low levels of alertness included 355 elderly riders (aged 46 to 65), 270 adult riders (aged 26 to 45), and 235 young riders (aged 12 to 25). In 2022, the number of motorcycles in Indonesia reached 125,305,332 units [28], and currently, the highest number of traffic accidents involves motorcycle riders. To reduce the risk of accidents among elderly drivers, further research is needed to determine the probability of single-vehicle accidents involving elderly motorcyclists, particularly in terms of human, road, and environmental factors. To date, no regulations in Indonesia set a maximum age limit for motorcycle riders. Additionally, existing studies have rarely addressed this issue. When discussed, researchers tend to recommend that older drivers use assistive tools or rely on support while riding, or that other road users adapt to older drivers to enhance safety [29].

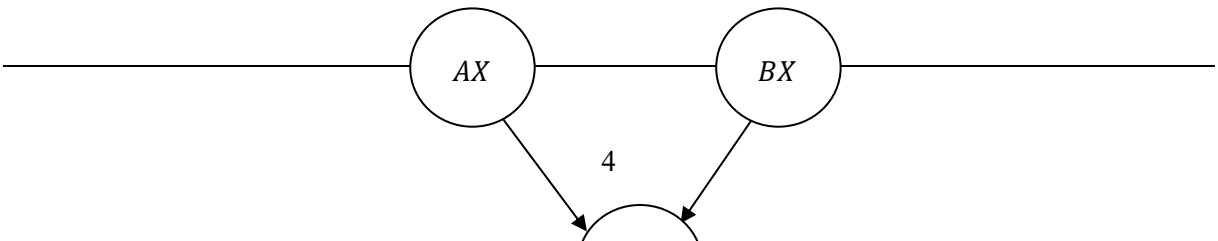
2. METHODS

The research location is in Riau Province, Indonesia. The reason for choosing this location is that the number of motorcycles in Riau Province is relatively high, with a total of 3,485,246 units. Riau is among the top 10 provinces in Indonesia with the highest number of motorcycles. According to police data from 2022, the number of road accident victims in Riau Province was 2,750 people, consisting of 651 fatalities, 357 people seriously injured, and 1,742 people with minor injuries [30]. This study involved 564 respondents: 425 respondents were used to analyze the accident probability model for elderly motorcycle riders, and 139 respondents were used to validate the model. The respondents' criteria included elderly riders aged 46 to 65 years who had experienced a motorcycle accident while riding. Respondents were selected randomly by asking candidates whether they had ever been involved in a motorcycle accident as a driver. If the candidate answered "no," the questioning was discontinued; if the candidate answered "yes," the questionnaire was completed in full. To develop this model, data were analyzed using Bayesian Network analysis with GeNie 2.0 software [31]. Since the data are probabilistic (uncertain), probabilistic analysis was carried out using GeNie 2.0 software, which is well-suited for this analysis. The author also conducted manual calculations, and the results were consistent with those produced by GeNie 2.0. However, manual calculations are slower and less efficient. Therefore, the author concluded that using GeNie 2.0 software is more effective for building the model. The formula used in this Bayesian Network analysis is shown in Eq. (1).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|-A)P(-A)} \tag{1}$$

Figure 1 and Figure 2 are examples of Bayesian Network structures. The Bayesian Network analysis in Figure 1 (which includes 3 variables) can be calculated using the following Eq. (2), where P is the probability.

$$P(Y) = P(Y|AX,BX) \times P(AX)P(BX) + P(Y|AX,-BX) \times P(AX)P(-BX) + P(Y|-AX,BX) \times P(-AX)P(BX) + P(Y|-AX,-BX) \times P(-AX)P(-BX) \tag{2}$$



https://doi.org/10.31436/iiumej.v26i3.3474

Figure 1. Example of Bayesian Network analysis with 3 variables.

The Bayesian Network analysis in Figure 2 (with 4 variables) can be calculated using the following Eq. (3):

$$P(Y) = P(Y|BX, CX, AX)P(BX|AX) \times P(CX|AX) + P(Y|BX, -CX, AX) \times P(-BX|AX) \times P(-CX|AX) + P(Y| - BX, -CX, AX) \times P(-BX|AX) \times P(CX|AX) + P(Y| - BX, -CX, AX) \times P(-BX|AX) \times P(-CX|AX) \quad (3)$$

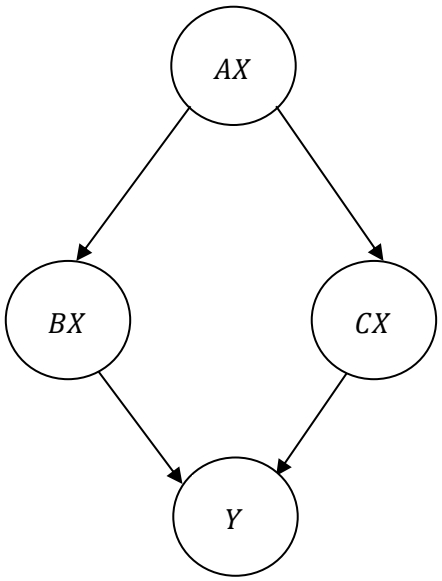


Figure 2. Example of analysis of a Bayesian network with 4variables.

Furthermore, to determine the accuracy of this model, it is first validated by calculating the Mean Absolute Deviation (MAD), which represents the average difference between the model's results and actual field conditions. The calculation utilizesEq. (4).

$$MAD = \frac{1}{N}\sum|Actual - Forecast| \quad (4)$$

The data used for model analysis and validation are not the same. For model validation, comparative data is used to determine the accuracy of the obtained model. The Mean Absolute Deviation (MAD) value is calculated to assess this accuracy, representing the average difference between the model's predicted values and the actual conditions in the field. The closer the MAD value is to zero, the more accurate the model. Suppose the model results closely match the field conditions. In that case, several scenarios are carried out, including: 1) Scenarios 1 and 2: the probability of a single-vehicle accident for elderly drivers based on the driver's gender; 2) Scenarios 3 and 4: the probability of a single-vehicle accident for elderly drivers based on the driver's stamina condition; 3) Scenarios 5 and 6: the probability of a single-

https://doi.org/10.31436/iiumej.v26i3.3474

vehicle accident for elderly drivers based on weather conditions; 4) Scenarios 7 and 8: the probability of a single-vehicle accident for elderly drivers based on road surface conditions. These scenarios were selected to examine the dominant influence of each variable that directly affects the probability of a single-vehicle accident, including the driver's gender (male and female drivers), the fatigue variable (fatigued and not fatigued), and the weather variable (driving in rainy and non-rainy conditions). The road surface condition variable (potholed and not potholed). These four variables influence the probability of a single-vehicle accident.

3. RESULTS AND DISCUSSION

The survey results show that 29% of elderly motorcyclists who had accidents had a sleep duration of 6 hours or less the night before the accident. Meanwhile, 27% of elderly motorcyclists experienced fatigue before the accident. Most elderly motorcyclists experienced accidents during short trips lasting 30 minutes or less. Furthermore, most of these accidents occurred between 12:00 and 18:00. Among elderly drivers who had accidents, 52% were male and 48% were female. The data also revealed that 59% of accidents occurred on flat, straight roads. Moreover, 38% of elderly drivers involved in accidents were driving under monotonous conditions, and 34% of the accidents happened on roads with potholes.

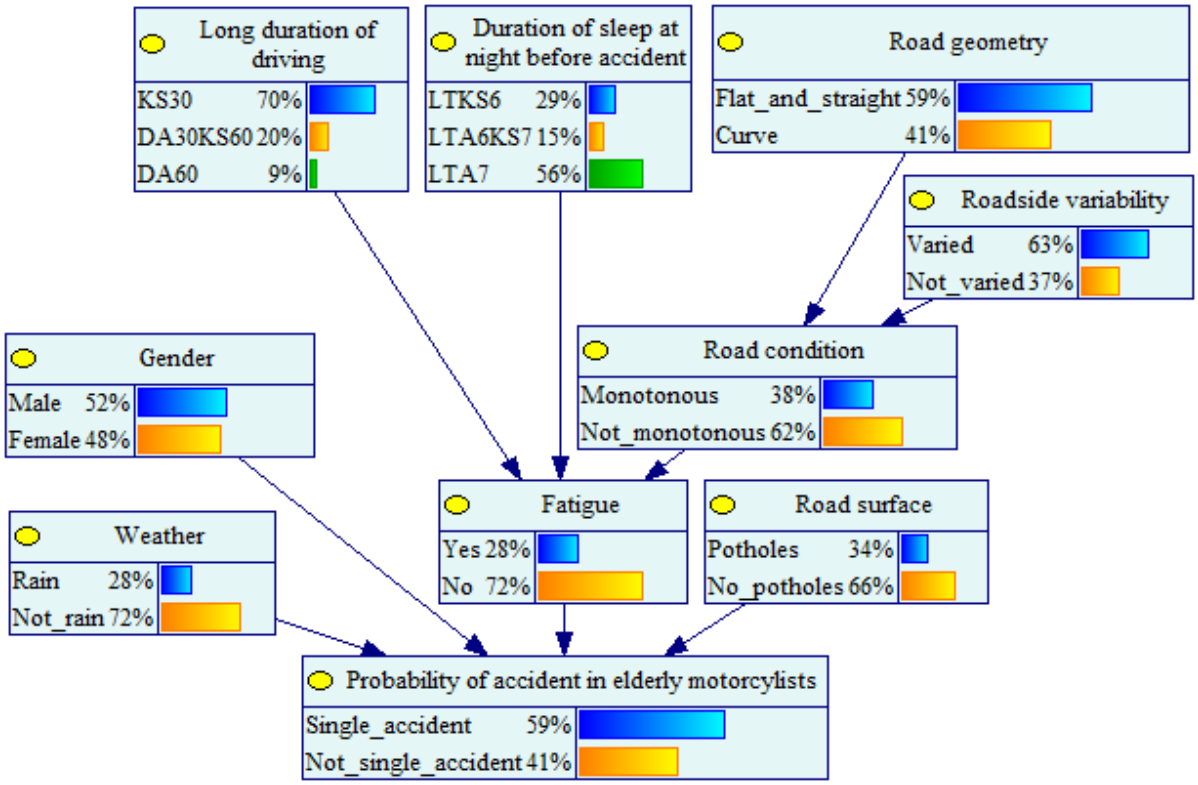


Figure 3. Structure model of the Bayesian Network showing the probability of a single-vehicle accident in elderly drivers.

Annotation:

- KS30 = Duration of driving for 30 minutes or less
- A30KS60 = Duration of driving for more than 30 minutes up to 60 minutes
- A60 = Duration of driving for more than 60 minutes
- LTKS6 = Sleep duration on the night before the accident of approximately 6 hours or less

https://doi.org/10.31436/iiumej.v26i3.3474

- LTA6KS7 = Sleep duration on the night before the accident of more than 6 hours up to 7 hours
- LTA7 = Sleep duration on the night before the accident of more than 7 hours

The results of the Bayesian Network analysis, conducted using GeNie 2.0 software, show that the probability of elderly motorcyclists experiencing a single-vehicle accident is 59%. In comparison, the likelihood of a non-single-vehicle accident is 41%. This indicates a high probability that elderly drivers will experience a single-vehicle accident. This is attributed to the increasing age of the driver, which results in a decrease in driving ability due to both physical and non-physical limitations, including limited vision, reduced stamina, decreased ability to concentrate, lack of familiarity with vehicle technology, and health problems, as shown in Figure 3. This study is consistent with research conducted by [1,4-11]. The equality model of the probability of a single-vehicle accident is shown in Table 1.

Table 1: Equality Model for the Probability of a Single-Vehicle Accident

P	G	E	W	RS	P(AC)
1	G1	E1	W1	RS1	P(AC)1=P(AC|G1,E1,W1,RS1,LD,DS,RG,RD,RC) P(E1|LD,DS) P(RC|RG,RV)
2	G1	E1	W1	RS2	P(AC)2=P(AC|G1,E1,W1,RS2,LD,DS,RG,RD,RC) P(E1|LD,DS) P(RC|RG,RV)
3	G1	E1	W2	RS1	P(AC)3=P(AC|G1,E1,W2,RS1,LD,DS,RG,RD,RC) P(E1|LD,DS) P(RC|RG,RV)
4	G1	E1	W2	RS2	P(AC)4=P(AC|G1,E1,W2,RS2,LD,DS,RG,RD,RC) P(E1|LD,DS) P(RC|RG,RV)
5	G1	E2	W1	RS1	P(AC)5=P(AC|G1,E2,W1,RS1,LD,DS,RG,RD,RC) P(E1|LD,DS) P(RC|RG,RV)
6	G1	E2	W1	RS2	P(AC)6=P(AC|G1,E2,W1,RS2,LD,DS,RG,RD,RC) P(E2|LD,DS) P(RC|RG,RV)
7	G1	E2	W2	RS1	P(AC)7=P(AC|G1,E2,W2,RS1,LD,DS,RG,RD,RC) P(E2|LD,DS) P(RC|RG,RV)
8	G1	E2	W2	RS2	P(AC)8=P(AC|G1,E2,W2,RS2,LD,DS,RG,RD,RC) P(E2|LD,DS) P(RC|RG,RV)
9	G2	E1	W1	RS1	P(AC)9=P(AC|G2,E1,W1,RS1,LD,DS,RG,RD,RC) P(E1|LD,DS) P(RC|RG,RV)
10	G2	E1	W1	RS2	P(AC)10=P(AC|G2,E1,W1,RS2,LD,DS,RG,RD,RC) P(E1|LD,DS) P(RC|RG,RV)
11	G2	E1	W2	RS1	P(AC)11=P(AC|G2,E1,W2,RS1,LD,DS,RG,RD,RC) P(E1|LD,DS) P(RC|RG,RV)
12	G2	E1	W2	RS2	P(AC)12=P(AC|G2,E1,W2,RS2,LD,DS,RG,RD,RC) P(E2|LD,DS) P(RC|RG,RV)
13	G2	E2	W1	RS1	P(AC)13=P(AC|G2,E2,W1,RS1,LD,DS,RG,RD,RC) P(E2|LD,DS) P(RC|RG,RV)
14	G2	E2	W1	RS2	P(AC)14=P(AC|G2,E2,W1,RS2,LD,DS,RG,RD,RC) P(E2|LD,DS) P(RC|RG,RV)
15	G2	E2	W2	RS1	P(AC)15=P(AC|G2,E2,W2,RS1,LD,DS,RG,RD,RC) P(E2|LD,DS) P(RC|RG,RV)
16	G2	E2	W2	RS2	P(AC)16=P(AC|G2,E2,W2,RS2,LD,DS,RG,RD,RC) P(E2|LD,DS) P(RC|RG,RV)
					∑ P(AC)

Annotation: P=Probability, G=Gender, G1=Male, G2=Female, E=Fatigue, E1=Yes, E2=No, W=Weather, W1=Rain, W2=Not rain, RS=Road surface, RS1=Potholes, RS2=Without potholes.

Furthermore, to determine the accuracy of the obtained model, the next step is to validate it. The validation results show that the MAD value, which is the average difference between the model results and real-world conditions, is 23%, as shown in Table 2. This means that the

https://doi.org/10.31436/iiumej.v26i3.3474

model has an accuracy of 77%. Since the level of model accuracy is relatively high, several scenarios can be created from the basic model to identify the dominant variables that influence the risk of single-vehicle accidents.

Table 2: Calculation of the MAD value

P	G	E	W	RS	Probability		Difference (%)
					Actual	Model	
1	G1	E1	W1	RS1	90,00	75,00	15,00
2	G1	E1	W1	RS2	66,67	83,00	16,33
3	G1	E1	W2	RS1	66,67	50,00	14,29
4	G1	E1	W2	RS2	40,91	53,00	13,87
5	G1	E2	W1	RS1	100,00	75,00	0,00
6	G1	E2	W1	RS2	85,71	58,00	25,33
7	G1	E2	W2	RS1	63,64	69,00	7,46
8	G1	E2	W2	RS2	34,21	54,00	20,67
9	G2	E1	W1	RS1	100,00	80,00	20,00
10	G2	E1	W1	RS2	0,00	100,00	100,00
11	G2	E1	W2	RS1	80,00	73,00	14,50
12	G2	E1	W2	RS2	37,50	56,00	13,14
13	G2	E2	W1	RS1	66,67	53,00	47,00
15	G2	E2	W2	RS1	40,00	57,00	23,67
16	G2	E2	W2	RS2	25,00	47,00	13,67
							23,00

Scenario 1 shows that elderly male motorcyclists have a 60% likelihood of experiencing a single-vehicle accident, as shown in Figure 4. In contrast, Scenario 2 shows that elderly female motorcyclists have a 59% likelihood of experiencing a single-vehicle accident, as shown in Figure 5. Male drivers tend to be more aggressive than female drivers. In addition, male drivers are more likely to underestimate road risks than female drivers. This means that the probability of a single-vehicle accident for elderly male and elderly female motorcyclists is almost the same and relatively high. This finding aligns with studies conducted by [19, 20].

Scenario 3 shows that elderly riders who drive while fatigued have a 65% likelihood of experiencing a single-vehicle accident, as shown in Figure 6. Meanwhile, Scenario 4 indicates that elderly riders who ride while not fatigued have a 57% likelihood of experiencing a single-vehicle accident, as shown in Figure 7. This suggests that the probability of a single-vehicle accident is higher for elderly motorcyclists who are fatigued than those who are not. However, the likelihood of a single-vehicle accident in both fatigued and non-fatigued conditions is above 50%. This suggests that elderly motorcyclists generally have poor stamina due to declining physical function. One of the limitations of elderly drivers is low stamina, which causes them to become fatigued quickly, especially during long trips. Fatigue can slow the driver's response time, increasing the risk of a single-vehicle accident. The results of this study are consistent with research conducted by [4, 32].

https://doi.org/10.31436/iiumej.v26i3.3474

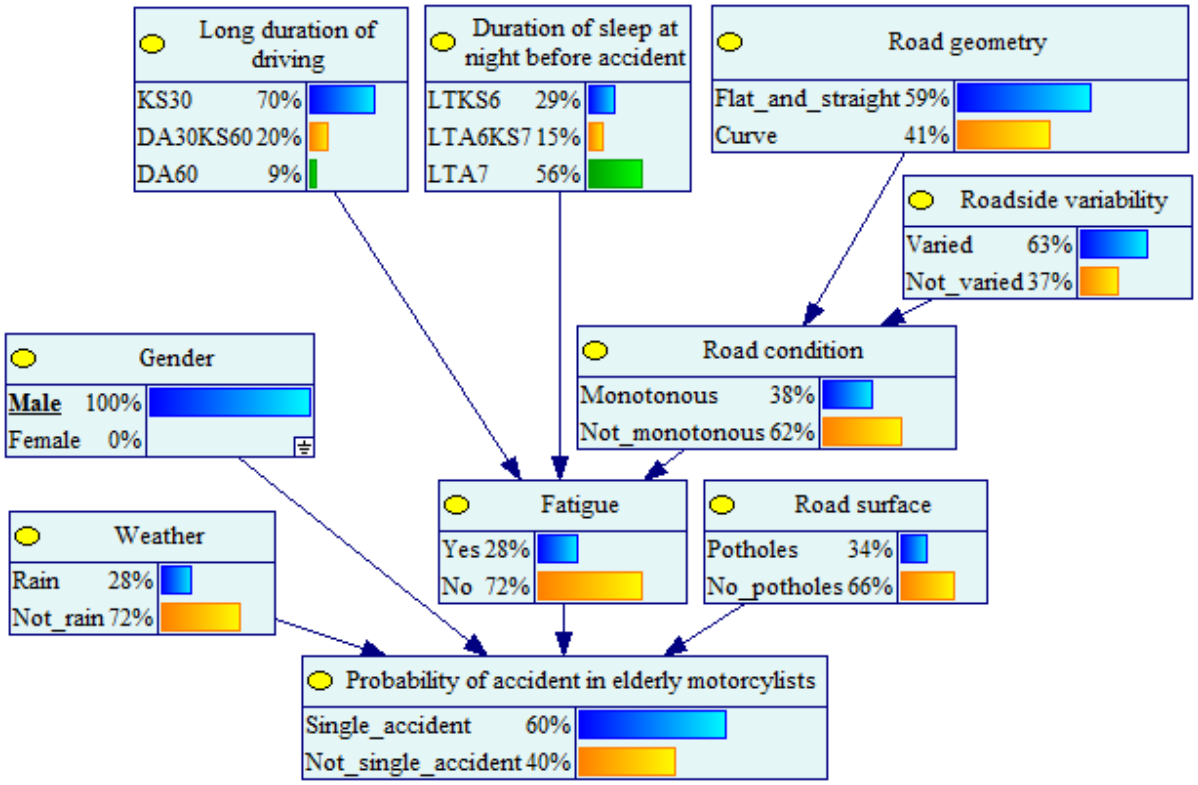


Figure 4. Scenario 1: Probability of a Single-Vehicle Accident Among Elderly Male Motorcyclists

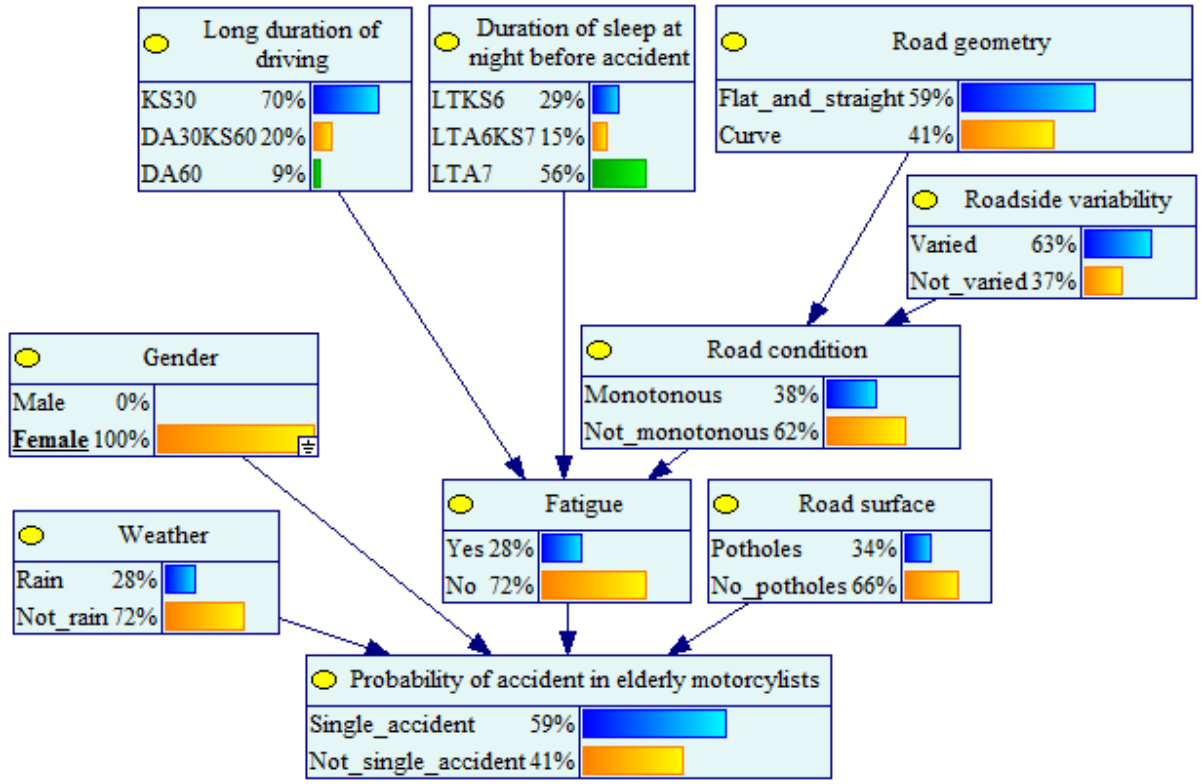


Figure 5. Scenario 2: Probability of a Single-Vehicle Accident among Elderly Female Motorcyclists

https://doi.org/10.31436/iiumej.v26i3.3474

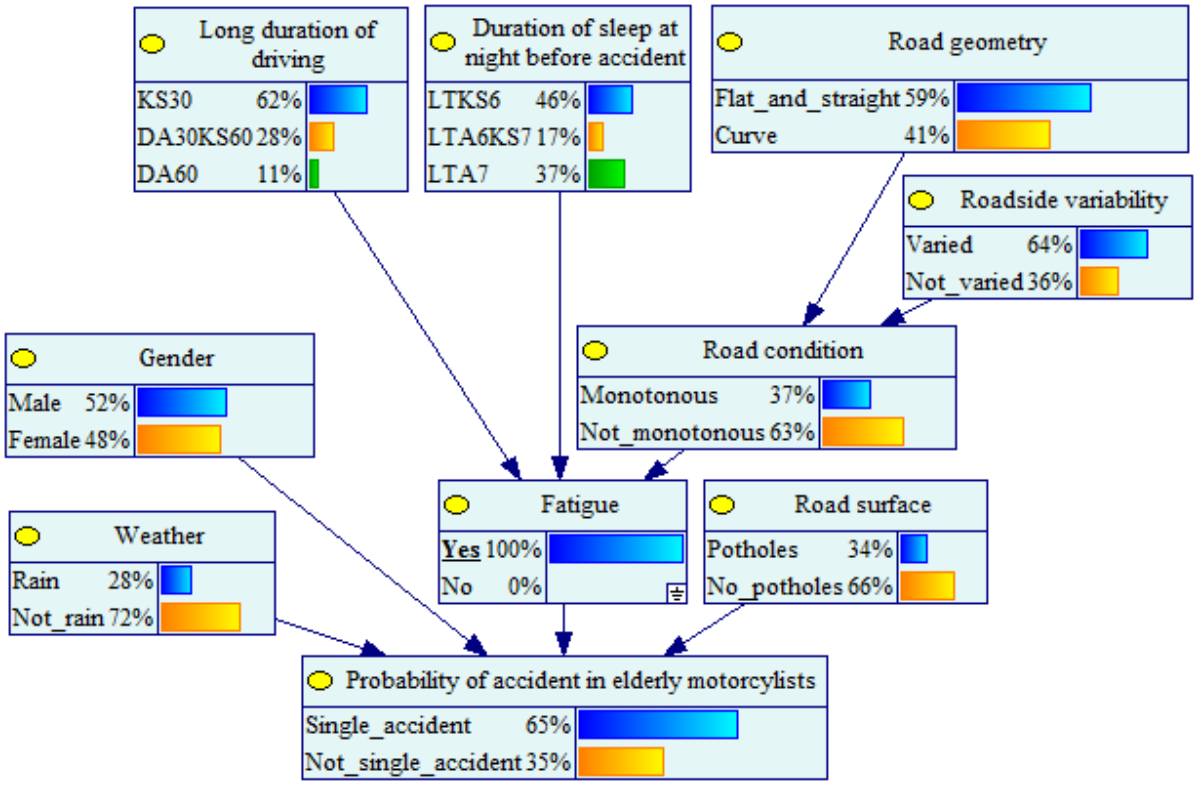


Figure 6. Scenario 3: Probability of a Single-Vehicle Accident Among Elderly Motorcyclists Driving While Fatigued

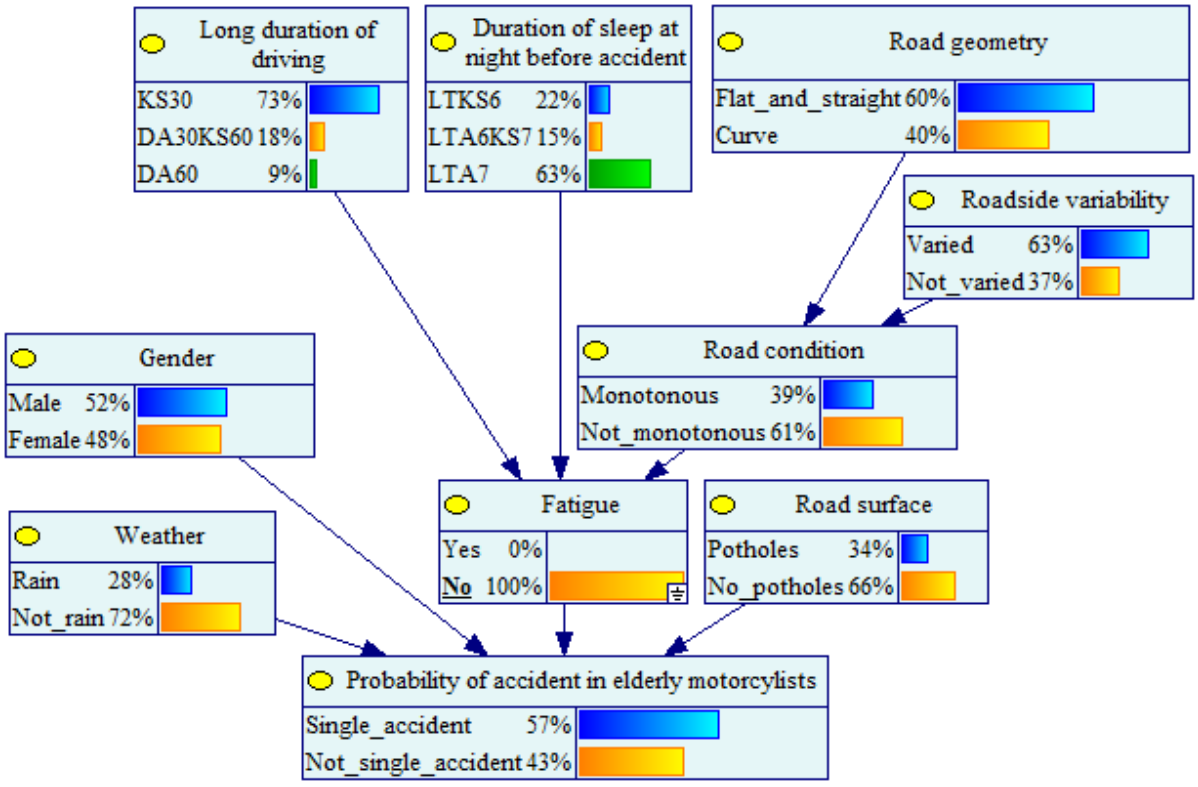


Figure 7. Scenario 4: Probability of a Single-Vehicle Accident Among Elderly Motorcyclists Driving While Not Fatigued.

https://doi.org/10.31436/iiumej.v26i3.3474

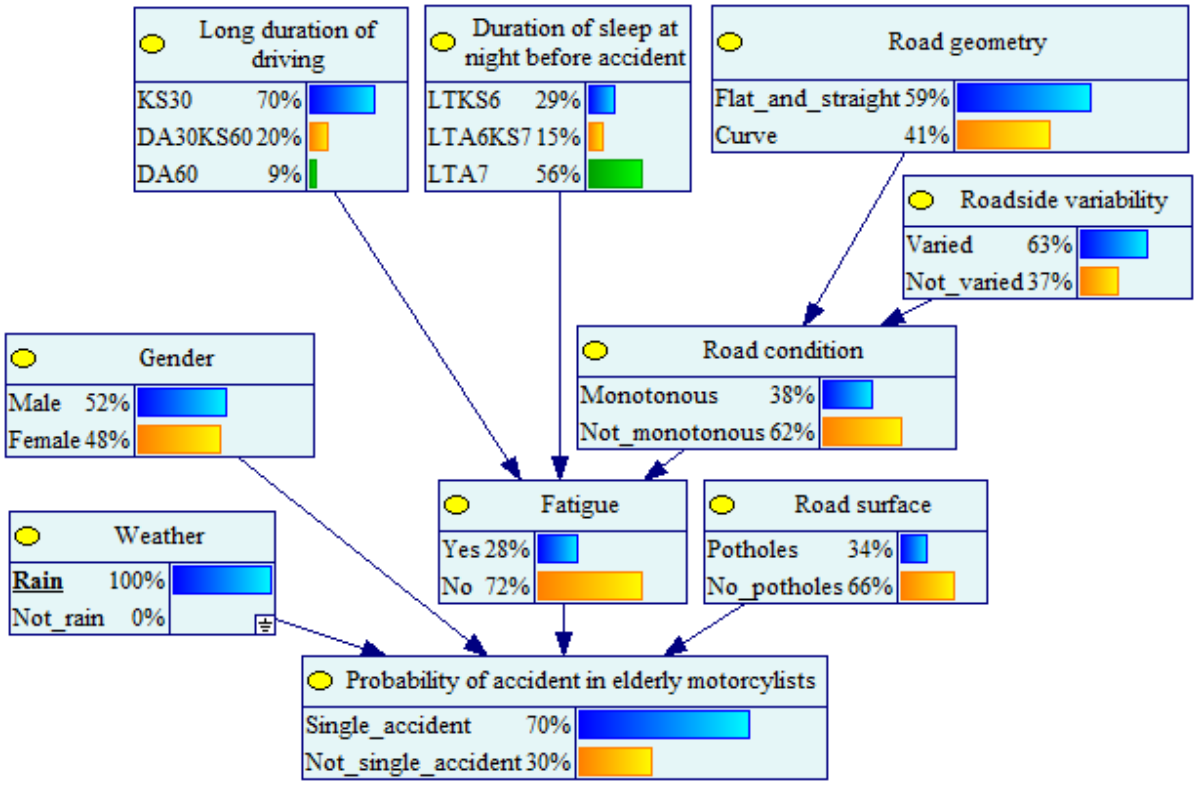


Figure 8. Scenario 5: Probability of a Single-Vehicle Accident Among Elderly Motorcyclists Driving in Rainy Conditions

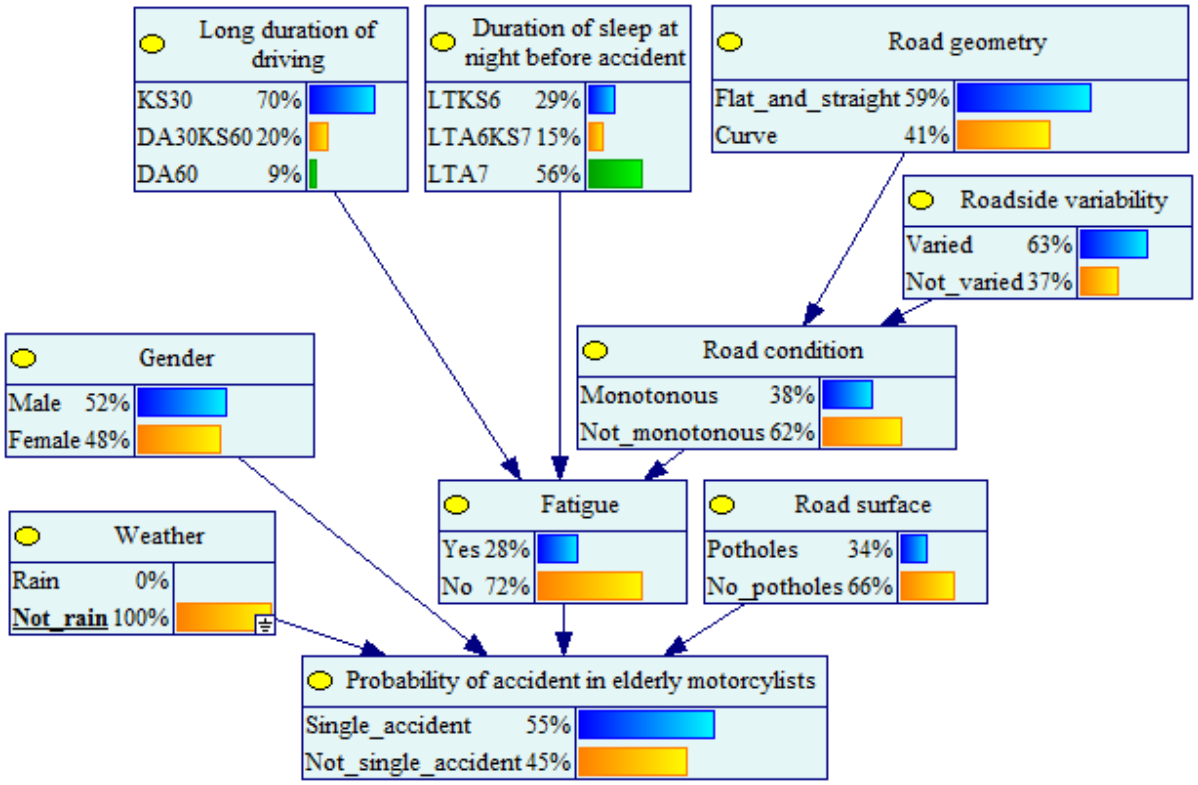


Figure 9. Scenario 6: Probability of a Single-Vehicle Accident Among Elderly Motorcyclists Driving in Non-Rainy Conditions

https://doi.org/10.31436/iiumej.v26i3.3474

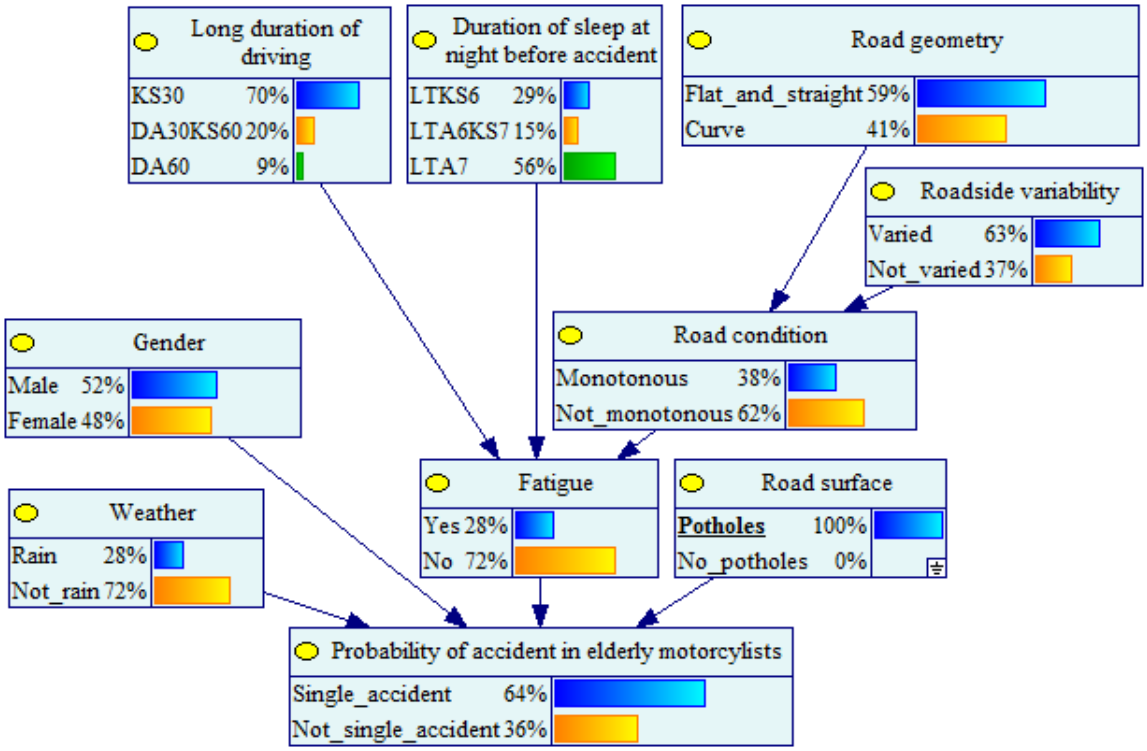


Figure 10. Scenario 7: Probability of a Single-Vehicle Accident Among Elderly Motorcyclists Driving on a Potholed Road.

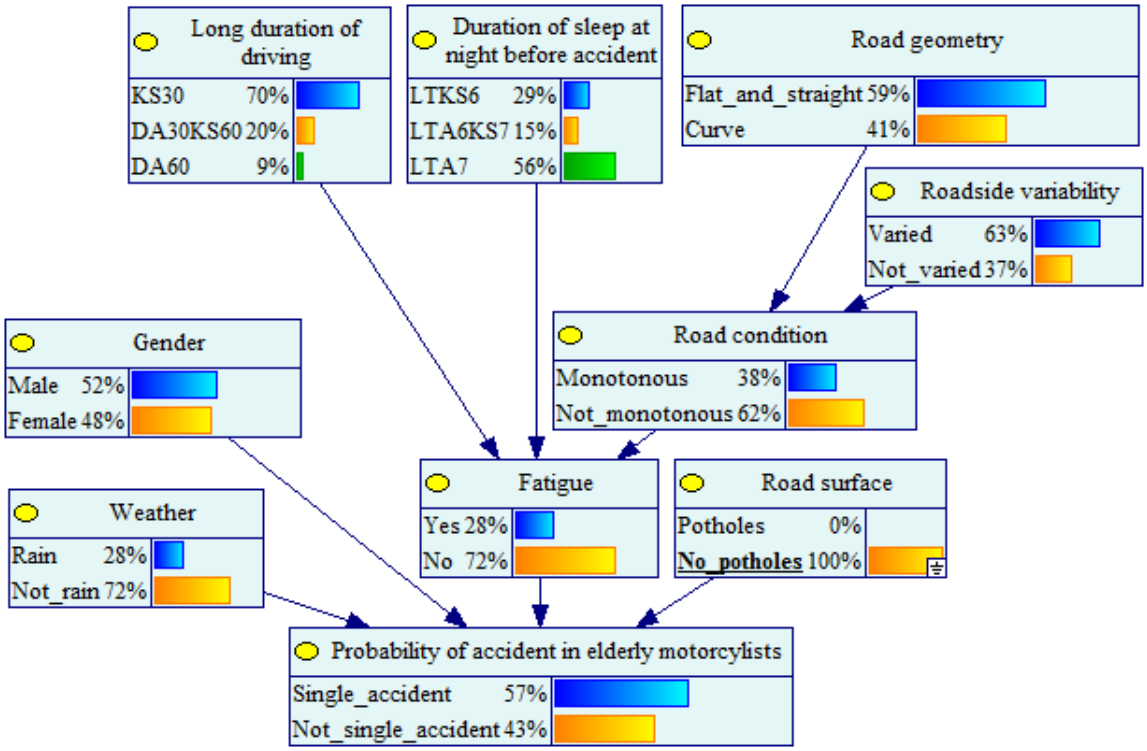


Figure 11. Scenario 8: Probability of a Single-Vehicle Accident Among Elderly Motorcyclists Driving on a Road Without Potholes

Scenario 5 indicates that elderly motorcyclists who ride in rainy conditions have a 70% likelihood of being involved in a single-vehicle accident, as illustrated in Figure 8. Meanwhile,

Scenario 6 shows that elderly motorcyclists who ride in non-rainy conditions have a 55% likelihood of experiencing a single-vehicle accident, as shown in Figure 9. Rainy conditions reduce rider visibility and create slippery road surfaces, increasing the risk of accidents. Under these circumstances, elderly drivers are more susceptible to single-vehicle accidents. These findings indicate that elderly motorcyclists riding in rainy conditions face a significantly higher risk of single-vehicle accidents. However, even in non-rainy conditions, the likelihood remains above 50%, suggesting a general decline in riding ability and skill among elderly individuals, primarily due to reduced physical function. As drivers age, their physical capabilities, such as vision, hearing, and stamina, tend to decline noticeably. Poor vision can prevent elderly drivers from recognizing traffic hazards ahead or around them. Similarly, the hearing ability of elderly drivers tends to decline, making it more challenging for them to detect warning signs, such as the sound of vehicle horns or engines. This condition increases their risk of accidents. Additionally, elderly drivers tend to become fatigued more quickly, further elevating the risk of accidents.

Scenario 7 shows that elderly motorcyclists riding on potholed roads are 64% likely to experience a single-vehicle accident, as shown in Figure 10. Meanwhile, Scenario 8 indicates that elderly motorcyclists riding on roads without potholes are 57% likely to experience a single-vehicle accident, as shown in Figure 11. This indicates that elderly motorcyclists on potholed roads are more likely to experience single-vehicle accidents than those without potholes. However, the probability of a single-vehicle accident for elderly motorcyclists remains above 50% in both scenarios. This suggests a decline in the ability and skills of elderly motorcyclists due to their physical and cognitive limitations. This decrease in ability is caused by elderly drivers' reduced ability to detect road conditions ahead or around them, which results in mistakes when responding to warnings or hazards nearby.

4. CONCLUSION

The probability of a single-vehicle accident among elderly drivers is relatively high, exceeding 50%. This indicates that elderly drivers are particularly vulnerable to such incidents. Based on eight scenarios analyzed in this study, several variables have been identified as significant contributors to single-vehicle accidents: male driver, driving while fatigued, driving in the rain, and driving on potholes. Among human factors, fatigue while driving is the most influential variable increasing the likelihood of a single-vehicle accident among elderly drivers. For environmental factors, driving in the rain is the primary contributor, while for road conditions, encountering potholes is a key factor contributing to single-vehicle accidents. These findings highlight the dominant factors contributing to single-vehicle accidents among elderly drivers, considering human, environmental, and road-related variables. This study is expected to serve as a reference for establishing a maximum age limit for motorcycle riders, as Indonesia currently lacks regulations regarding the maximum age for riding a motorcycle.

REFERENCES

[1] Bucsuházy K, Matuchová E, Zůvala R, Moravcová P, Kostíková M, Mikulec R. (2020) Human factors contributing to the road traffic accident occurrence.Transportation Research Procedia, 45(2019), 555–561.https://doi.org/10.1016/j.trpro.2020.03.057

[2] Chouhan SS, Kathuria A, Sekhar CR. (2023) The motorcycle rider behaviour questionnaire as a predictor of crashes: A systematic review and meta-analysis. IATSS Research, 47 (2023):61–72

[3] Lumba P, Ariyanto A, Fathoni A. (2024)Strategies for Enhancing Traffic Safety among Adolescent and Adult Motorcycle Riders in Indonesia. Jordan Journal of Civil Engineering, Volume 18(2):293-307

[4] Chu H-C. (2020) Risky behaviors of older taxi drivers and suggested requirements for renewing their professional driver's licenses. Transportation Research Interdisciplinary Perspectives, 8 (2020), 100272. https://doi.org/10.1016/j.trip.2020.100272

[5] Yeung JS, Wong YD. (2015)Effects of driver age and experience in abruptonset hazards. Accident Analysis and Prevention, 78(2015):110–117. http://dx.doi.org/10.1016/j.aap.2015.02.024

[6] Wada S, Hagiwara T, Hamaoka H, Ninomiya Y, Ohiro T, Tada M. (2020) Differences in situational awareness between elderly and middle-aged drivers in level 2 automated vehicles versus nonautomated vehicles. Asian Transport Studies, 6 (2020), 100014.https://doi.org/10.1016/j.eastsj.2020.100014

[7] Classen S, Mason J, Hwangbo SW, Wersal J, Rogers J, Sisiopiku V. (2021). Older drivers' experiencewith automated vehicle technology. Journal of Transport & Health, 22 (2021), 101107.https://doi.org/10.1016/j.jth.2021.101107

[8] Stamatiadis N, Kirk A. (2019) Use of Technology-Based Strategies for Older Driver Risk Mitigation. AIIT 2nd International Congress on Transport Infrastructure and Systems in a Changing World. (TIS ROMA 2019), 23rd-24th September 2019, Rome, Italy. Transportation Research Procedia, 45(2020): 651–658. https://doi.org/10.1016/j.trpro.2020.03.045

[9] Nguyen H, Coxon K, Brown J, Neville N, Tanna GLD, Hsieh Y-W, Keay L. (2023) Older drivers in Australia and advanced vehicle technologies: What are their opinions? A qualitative study. Journal of Transport & Health, 31 (2023) 101646

[10] Toepper M, Austerschmidt KL, Schlueter DA, Koenig J, Beblo T, Driessen M. (2024) On-road driving performances at traffic signs and signals, complex intersections and left turns distinguish fit and unfit older drivers. Transportation Research Part F: Psychology and Behaviour, 102(2024):54–63

[11] Skyving M, Forsman A, Willstrand TD, Laflamme L, Moller J. (2021) Medical impairment and road traffic crashes among older drivers in Sweden – A national, population-based, case-control study. Accident Analysis and Prevention, 163 (2021), 106434. https://doi.org/10.1016/j.aap.2021.106434

[12] Madvari RF, Sefidkar R, Halvani GH, Alizadeh HM. (2023) Quantitative indicators of street lighting with mood, fatigue, mental workload and sleepiness in car drivers: Using generalized structural equation modeling. Heliyon, 9 (2023) e12904. https://doi.org/10.1016/j.heliyon.2023.e12904

[13] Davidovic, J., Pešic, D., Lipovac, K., & Antic, B. (2020). The Significance of developing road safety performance indicators related to Driver Fatigue. AIIT 2nd International Congress on Transport Infrastructure and Systems in a changing world (TIS ROMA 2019), 23rd-24th September 2019, Rome, Italy. Transportation Research Procedia, 45 (2020): 333–342. https://doi.org/10.1016/j.trpro.2020.03.024

[14] Lumba P, Edison B, Fahmi K, Sibarani AS, Ariyanto A, Hidayat A, Rahmi A, Rismalinda.(2022) Effects of Sleep Deprivation on Probability of Traffic Violations in Motorcyclists; Analysis Using Bayesian Network. Science and Technology Asia, Vol.27 No.2 April - June 2022, Page [115-125]. doi: 10.14456/scitechasia.2022.29.

[15] Elvik R. (2023) Driver mileage and accident involvement: A synthesis of evidence. Accident Analysis and Prevention, 179 (2023), 106899. https://doi.org/10.1016/j.aap.2022.106899

[16] Lumba P, Ariyanto A, Fathoni A. (2024a). "STRATEGIES TO REDUCE THE NUMBER OF SEVERELY INJURED VICTIMS IN ADOLESCENT MOTORCYCLE RIDERS". IIUM Engineering Journal, Vol. 25(1):153-166.https://doi.org/10.31436/iiumej.v25i1.2997

[17] Wijayanto T, Marcillia SR, Lufityanto G, Wisnugraha BB, Alma TG, Abdianto RU. (2021) The effect of situation awareness on driving performance in young sleep-deprived drivers. IATSS Research, 45 (2021): 218–225. https://doi.org/10.1016/j.iatssr.2020.10.002

[18] Takeyama E, Tomooka K, Wada H, Sato S, Sakiyama N, Shirahama R, Tanigawa T. (2023) Association between daytime sleepiness and motor vehicle accidents among Japanese male taxi drivers. IATSS Research, 47(2023):299–304. https://doi.org/10.1016/j.iatssr.2023.05.002

[19] Kim K, Matsuhashi K, Ishikawa M. (2023) Analysis of primary-party traffic accident rates per driver in Japan from 1995 to 2015: Do older drivers cause more accidents?. IATSS Research, 47(2023):447–454

[20] Dzinyela R, Adanu EK, Lord D, Islam S. (2023) Analysis of factors that influence injury severity of single and multivehicle crashes involving at-fault older drivers: A random parameters logit with heterogeneity in means and variances approach. Transportation Research Interdisciplinary Perspectives, 22 (2023), 100974

[21] Ulleberg P, Bjørnskau T, Fostervold KI. (2022) Does age matter? Examining age-dependent differences in at-fault collisions after attending a refresher course for older drivers. Transportation Research Part F: Psychology and Behaviour, 87 (2022): 379–390. https://doi.org/10.1016/j.trf.2022.04.016

[22] Keskinen E. (2014). Education for older drivers in the future. IATSS Research, 38(2014):14–21

[23] Zhao L, Wang C, Yang H, Wu X, Zhu T, Wang J. (2023) Exploring injury severity of Non-Motor vehicle riders involving in traffic accidents using the generalized ordered logit model. Ain Shams Engineering Journal, 14 (2023), 101962. https://doi.org/10.1016/j.asej.2022.101962

[24] Larue GS, Rakotonirainya A, Pettitt AN. (2011) Driving performance impairments due to hypovigilance on monotonous roads. Accident Analysis and Prevention, 43(2011):2037– 2046, Elsevier. https://doi.org/10.1016/j.aap.2011.05.023

[25] Regev S, Rolison JJ, Moutari S. (2018). "Crash risk by driver age, gender, and time of day using a new exposure methodology". Journal of Safety Research, 66 (2018), 131–140. https://doi.org/10.1016/j.jsr.2018.07.002

[26] Ministry of Transportation. (2020). Available from: https://dephub.go.id/post/read/korban-kecelakaan-lalin-didominasi-usia-produktif,-menhub-ajak-para-pelajar-selalu-disiplin-berlalu-lintas-dan-utamakan-aspek-keselamatan

[27] Zhu Y, Jiang M, Yamamoto T. (2022)Analysis on the driving behavior of old drivers by driving recorder GPS trajectory data. Asian Transport Studies, 8 (2022), 100063. https://doi.org/10.1016/j.eastsj.2022.100063

[28] BPS-STATISTIC INDONESIA. Land Transportation Statistics. 2023. Available from:https://www.bps.go.id/en/publication/2024/11/25/cdcf9b5e74dd2e9bb3458ee4/land-transportation-statistics-2023.html

[29] Haan TD, Stuiver A, Lorist MM, Waard DD. (2022) Other road users' adaptations to increase safety in response to older drivers' behavior. Transportation Research Part F: Psychology and Behaviour, 84 (2022): 277–286. https://doi.org/10.1016/j.trf.2021.12.009

[30] Polda Provinsi Riau; 2022. Available from: https://www.detik.com/sumut/hukum-dan-kriminal/d-6490638/polda-ada-651-orang-tewas-kecelakaan-di-riau-sepanjang-2022

[31] Software GeNie. (2023). Available from: https://download.bayesfusion.com/files.html?category=Academia

[32] Abdubrani R, Mustafa M, Zahari Zl. (2023). A robust framework for driver fatigue detection from EEG signals using enhancement of modified z-score and multiple machine learning architectures. IIUM Engineering Journal, 24(2):354372.https://doi.org/10.31436/iiumej.v24i2.2799

https://doi.org/10.31436/iiumej.v26i3.3606

CORROSION EVALUATION ON FLY ASH AND SILICA FUME CONCRETE USING NDT METHODS

AHMAD ZAKI[1,2*], FERDI ARDIAN SYAH[1], SRI ATMAJA P. ROSYIDI[1]
KHARISMA WIRA NINDHITA[2], ZAINAH IBRAHIM[3]

[1]*Department of Civil Engineering, Faculty of Engineering, Universitas Muhammadiyah Yogyakarta, Indonesia*
[2]*Magister of Civil Engineering, Faculty of Engineering, Universitas Muhammadiyah Yogyakarta, Indonesia*
[3]*Department of Civil Engineering, Faculty of Engineering, University of Malaya, Malaysia*

**Corresponding author: ahmad.zaki@umy.ac.id*

(Received: 3 February 2025; Accepted: 7 March 2025; Published online: 9 September 2025)

ABSTRACT: Corrosion of reinforced concrete is a major problem that can weaken a structure and even cause it to fail. On a macro level, this affects the safety and durability of important structures like bridges, buildings, and roads, leading to high repair costs and safety risks. On a micro level, corrosion begins with chemical reactions between steel and concrete, resulting in cracks, surface damage, and weakening of the concrete, which in turn reduces its strength and performance. This study examines the effects of using industrial waste, namely fly ash (FA) and silica fume (SF), as alternative solutions to improve reinforced concrete's corrosion resistance and mechanical performance. Concrete is made with FA substitutions of 10%, 20%, and 30% of the cement weight, and SF of 5%, 10%, and 15%, respectively. Compressive strength and flexural tests were performed to evaluate the mechanical properties, and corrosion analysis was conducted through non-destructive testing (NDT) to assess the corrosion resistance. Corrosion testing was carried out with accelerated corrosion for 48, 96, and 168 hours. Microstructural analysis using SEM was also performed to observe the effect of corrosion on the concrete. The results showed that the compressive strength of normal concrete and concrete with FA and SF decreased from 37 MPa to 25-35 MPa, respectively. The flexural strength of normal concrete and concrete with FA and SF decreased from 13 to 2 MPa due to increased corrosion rates, respectively. However, the normal concrete's resistivity value and impact echo frequency were lower than those of FA- and SF-mixed concrete. Significant resistivity values and frequency reductions were observed after corrosion, with V3 specimens (30% FA and 15% SF) showing the best corrosion durability performance.

ABSTRAK: Hakisan konkrit bertetulang merupakan masalah utama yang boleh melemahkan struktur dan bahkan menyebabkan kegagalan keseluruhan. Pada peringkat makro, ia menjejaskan keselamatan dan ketahanan struktur penting seperti jambatan, bangunan, dan jalan raya, sekali gus membawa kepada kos pembaikan yang tinggi serta risiko keselamatan. Pada peringkat mikro, hakisan bermula dengan tindak balas kimia antara keluli dan konkrit, menyebabkan rekahan, kerosakan permukaan, dan kelemahan pada konkrit, yang akhirnya mengurangkan kekuatan dan prestasinya. Kajian ini meneliti kesan penggunaan sisa industri iaitu abu terbang (FA) dan wap silika (SF) sebagai penyelesaian alternatif untuk meningkatkan rintangan hakisan dan prestasi mekanikal konkrit bertetulang. Konkrit dihasilkan dengan penggantian FA sebanyak 10%, 20%, dan 30% daripada berat simen serta SF sebanyak 5%, 10%, dan 15%. Ujian kekuatan mampatan dan lenturan dijalankan bagi menilai sifat mekanikal, manakala analisis hakisan melalui kaedah tanpa musnah (NDT) digunakan untuk menilai rintangan hakisan. Ujian hakisan dijalankan menggunakan kaedah

https://doi.org/10.31436/iiumej.v26i3.3606

hakisan dipercepat selama 48, 96, dan 168 jam. Analisis mikrostruktur menggunakan SEM turut dilakukan bagi memerhati kesan hakisan terhadap konkrit. Hasil kajian menunjukkan bahawa kekuatan mampatan konkrit normal serta konkrit dengan FA dan SF menurun daripada 37 MPa kepada 25–35 MPa. Kekuatan lenturan konkrit normal serta konkrit dengan FA dan SF juga menurun daripada 13 MPa kepada 2 MPa apabila kadar hakisan meningkat. Namun begitu, nilai rintangan dan frekuensi gema hentakan bagi konkrit normal adalah lebih rendah berbanding konkrit yang dicampur FA dan SF. Nilai rintangan serta pengurangan frekuensi yang ketara diperhatikan selepas hakisan, dengan spesimen V3 (30% FA dan 15% SF) menunjukkan prestasi ketahanan hakisan yang terbaik.

KEYWORDS: *Fly Ash, Silica Fume, Concrete, Corrosion, NDT*

INTRODUCTION

Corrosion of reinforcements is one of the main causes of premature damage to reinforced-concrete (RC) structures. The corrosion process is caused by the reaction between reinforcing steel and hydroxide (OH) ions derived from the reaction between oxygen and water. This can occur because of damage to the passive layer of the rebar, known as depassivation, which is caused by the infiltration of chloride ions (Cl^-) into concrete [1-3]. Steel reinforcement has a corrosion-protective layer known as a passive film. However, this layer can be destroyed by the chloride ions or carbon reactions around the reinforcement. When the concentration of chloride ions reaches a critical threshold on the surface of the steel reinforcement, the coating passivity film is eroded, and the corrosion process occurs immediately [1]. To overcome this corrosion problem, the development and use of mixed cement has begun to develop rapidly in the construction industry, mainly owing to the consideration of cost savings, energy savings, environmental protection, and resource conservation. Fly ash, a silica material obtained from various thermal power plants, is considered a cement material for concrete [4]. In general, there are two types of additives: mineral and chemical. The better the characteristics of the additives used, the better the concrete will be produced.

Fly ash (FA) is a very fine coal combustion residue. Fly ash has similar physical properties to cement, namely, the fineness of the grains and the chemical content in the form of silica [5]. The second mineral additive is Silica Fume (SF). Silica fume is a very fine pozzolan material composed mostly of amorphous silica produced by electric arc furnaces as a byproduct of silicon production. Silica fumes contain silica compounds and have very fine grains up to 100 times smaller than cement grains. Saraswathy et al. [4] examined the effects of activated fly ash on concrete's corrosion resistance and strength. Concrete with 10%, 20%, 30%, and 40% activated FA (chemical and thermal activation) was tested for compressive strength at 7, 14, 28, and 90 days. The results showed that chemical activation was more effective in improving the corrosion resistance and strength of concrete, with the replacement of 20%-30% of activated FA improving the corrosion resistance of concrete. Zaki et al. [2] examined the effect of corrosion on palm shell concrete and mask fibers using non-destructive testing (NDT), i.e., resistivity and impact echo methods. Concrete underwent accelerated corrosion with a 5% NaCl solution and a DC power supply. The results showed that corrosion decreases the strength and durability of concrete, and the NDT method effectively detects corrosion in concrete.

Several studies have shown that the addition of pozzolan additives to concrete, such as FA and SF, increases the sealing properties against water penetration. The use of FA can modify the characteristics of fresh and hard concrete by improving the fatigue level, strength, and abrasion resistance [6]. According to Saraswathy et al. [4], a concrete mix that uses FA and SF as a substitute material for cement produces good concrete strength compared to normal concrete. Despite growing interest in sustainable concrete technologies, limited research has

https://doi.org/10.31436/iiumej.v26i3.3606

been conducted on evaluating corroded concrete incorporating supplementary cementitious materials (SCMs) like FA and SF, mainly through NDT methods. Most existing studies focus on the mechanical strength development of FA/SF concrete [5, 6], corrosion resistance improvements due to pozzolanic reactions [4, 5], chloride permeability and durability metrics using destructive tests (DT) [2, 5], and NDT methods (i.e., ultrasonic pulse velocity (UPV), rebound hammer, ground-penetrating radar (GPR), electrical resistivity) on ordinary Portland cement (OPC)-based concrete, often without incorporating SCMs.

This study achieved the maximum concrete strength by replacing 20%-30% of the cement. Therefore, it is essential to conduct further research with cement replacement of fly ash by 10%, 20%, and 30%, and silica fume by 5%, 10%, and 15%, respectively. Variations were selected based on several references' optimum FA and SF values. Thus, the strength ratio between normal concrete and mixed concrete improves when the composition of fly ash and silica fume improves. In addition, this research also reviews corrosion detection using the NDT method with resistivity and impact echo tools, which can be used as reference parameters when checking structures made from fly ash and silica fume, especially before and after corrosion. This research also examines the influence of corrosion on the mechanical properties of concrete mixed with fly ash and silica fume.

RESEARCH METHODS

2.1. Materials

This research was conducted at the Civil Engineering Construction Materials Laboratory, Universitas Muhammadiyah Yogyakarta. The fine aggregate material used in this study originated from the Progo River, whereas the coarse aggregate was from Clereng, Yogyakarta. The test specimen was made using Portland Composite Cement (PCC) from Dynamix with Ø12 mm reinforced iron with a length of 60 cm. The mold of the specimen was 50 cm × 10 cm × 10 cm, according to the size of the specimen to be made. The fly ash material was obtained from Cilacap PLTU, whereas the silica fume was obtained from PT. Sika Indonesia. The laboratory test results are presented in Table 1. Figure 1 shows the sieve analysis chart.

Table 1. Fine and coarse aggregate test data

Types of Testing	Fine Aggregate Value	Coarse Aggregate Value	Unit
Fine modulus of grain	2.435	-	-
Bulk specific gravity	1.837	2.642	-
The specific gravity of saturated dry	2.243	2.707	-
Specific gravity apparent	3.094	2.826	-
Water absorption	22.16	2.46	%
Sludge rate	1.00	1.00	%
Abrasion	-	15.75	%

https://doi.org/10.31436/iiumej.v26i3.3606

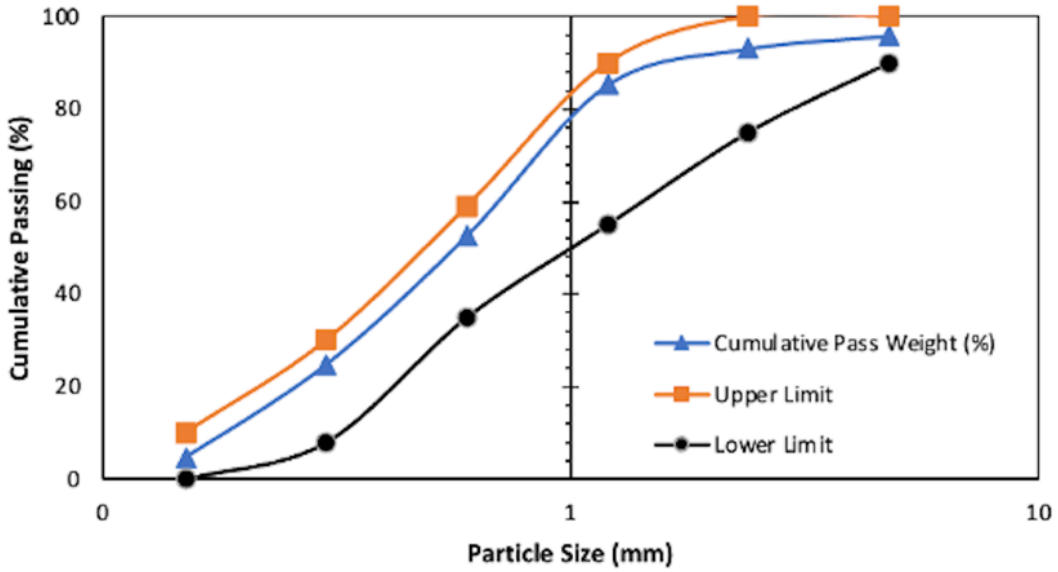


Figure 1. Cumulative passing (%)

2.2. Mix Design Concrete

Concrete mix planning using ACI 211.1-91 concerns procedures for maintaining proportions for normal concrete production with a planned concrete quality of 30 MPa [7]. Proportions of fly ash (FA) of 10%, 20%, and 30%, and silica fume (SF) of 5%, 10%, and 15% were added, respectively. Fly ash proportions of 10%, 20%, and 30% and silica fume proportions of 5%, 10%, and 15% were selected to evaluate their individual systematic and combined effects on concrete properties, including strength, durability, and workability. These ranges fall within standard practice and allow for optimizing the pozzolanic reaction, cost-effectiveness, and sustainability while balancing early and long-term performance. The mix design proportions per m³ are shown in Table 2. NC is normal concrete, V1 is concrete with 10% FA and 5% SF, V2 is concrete with 20% FA and 10% SF, and V3 is concrete with 30% FA and 15% SF.

Table 2. Proportion mix design per m^3

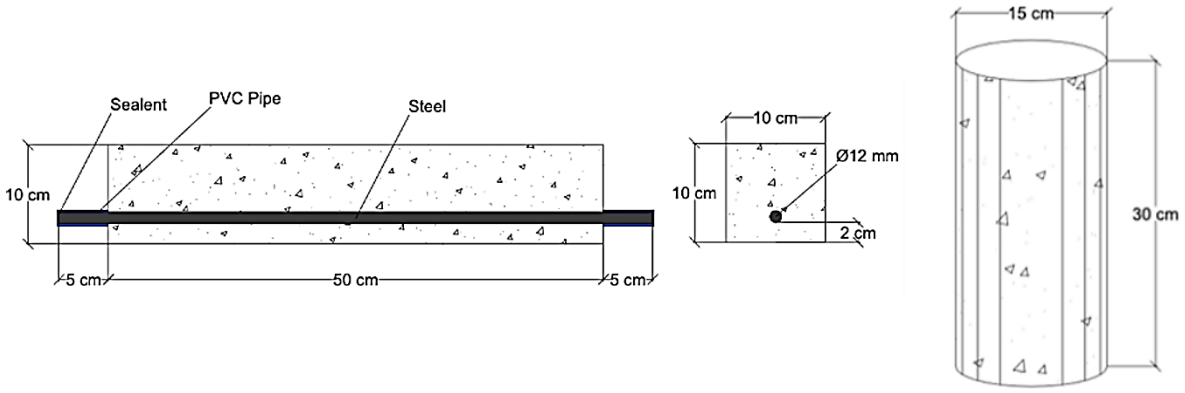
Material	TOTAL				Unit
	NC	V1	V2	V3	
Water	205	205	205	205	liter
Cement	484.21	411.44	338.84	266.23	kg
Sand	641.52	641.52	641.52	641.52	kg
Gravel	984.75	984.75	984.75	984.75	kg
Fly Ash	-	48.40	96.81	145.21	kg
Silica Fume	-	24.20	48.40	72.61	kg

2.3. Specimens Fabrication

The test specimen was in a beam (50 × 10 × 10 cm) and cylindrical form (30 × 15 cm), as shown in Figure 2. The main reinforcement size was Ø12 mm. A small cable was attached to the end of the reinforcement as a channel for accelerating concrete corrosion using a DC power supply. The test specimen was made based on a mixture proportion of fly ash (10%, 20%, and 30%) and silica fume (5%, 10%, and 15%). After the slump test, the concrete mixture was

https://doi.org/10.31436/iiumej.v26i3.3606

poured into the formwork until it was full and dense so air cavities did not form. The concrete was left to harden for 24 hours and then removed from the formwork.



(a) (b)

Figure 2. Specimen schematic (a) beam specimen and (b) cylinder specimen

This test was performed based on ACI 211.1-91 [7] to determine the concrete's viscosity. The slump test results were expressed in centimeters using an Abrams cone tool, a fist stick, and a meter. This study was planned to have a slump value of 25-100 mm. The test result slumps are shown in Table 3. Curing in concrete is a method used to maintain temperature stability, prevent moisture changes inside and outside the concrete, and help speed up the hydration process of the concrete. In this study, curing was performed by soaking the specimens for 28 days.

Table 3. Test scores slump

	Variations		Slump Value (cm)
	Fly Ash (%)	Silica Fume (%)	
NC	0	0	10
V1	10	5	9.5
V2	20	10	8
V3	30	15	7

2.4. Resistivity Testing

The resistivity test was performed when the specimen was 28 days old. The test was performed before and after the corrosion acceleration process to compare the test results based on AASHTO TP95-14 [8]. To obtain more complete and accurate data, the resistivity measurement tool employs the four-point probe method, which is performed three times at each point, ensuring that the data obtained is comprehensive and accurate. The details of the point division in which the resistivity test was carried out are shown in Figure 3. The sides are divided into four areas: the upper side, directly connected by the cable; the lower side; the right side; and the left side. Point division is used to obtain accurate and even results, so the average is obtained.

https://doi.org/10.31436/iiumej.v26i3.3606

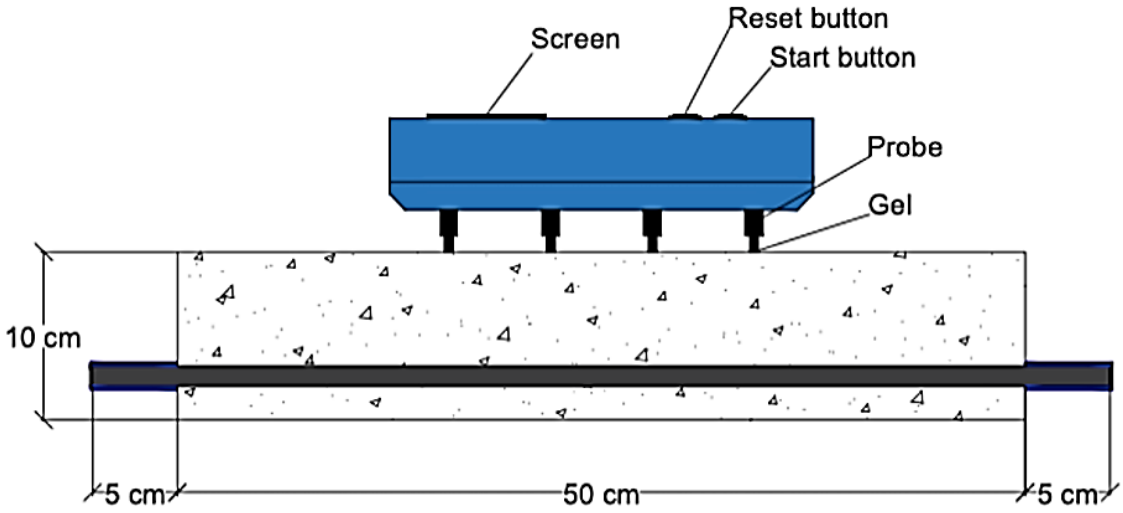


Figure 3. Testing illustration of resistivity

2.5. Impact Echo Testing

Impact-echo testing was performed when the specimen was 28 days old. The test was performed before and after the corrosion acceleration process to compare the test results. The impact-echo method was used to detect defects in concrete based on ASTM C1383 [9]. This is based on monitoring surface movements due to short-term mechanical impacts. The concrete is tapped several times using an impact echo test tool, and the sensor receives signals at a predetermined distance between it and the point of impact. Figure 4 illustrates the impact echo test. Impact echo testing was performed with two sensors at 5, 10, 15, and 20 cm hit points. The details of the impact echo test at each distance are shown in Figure 4.

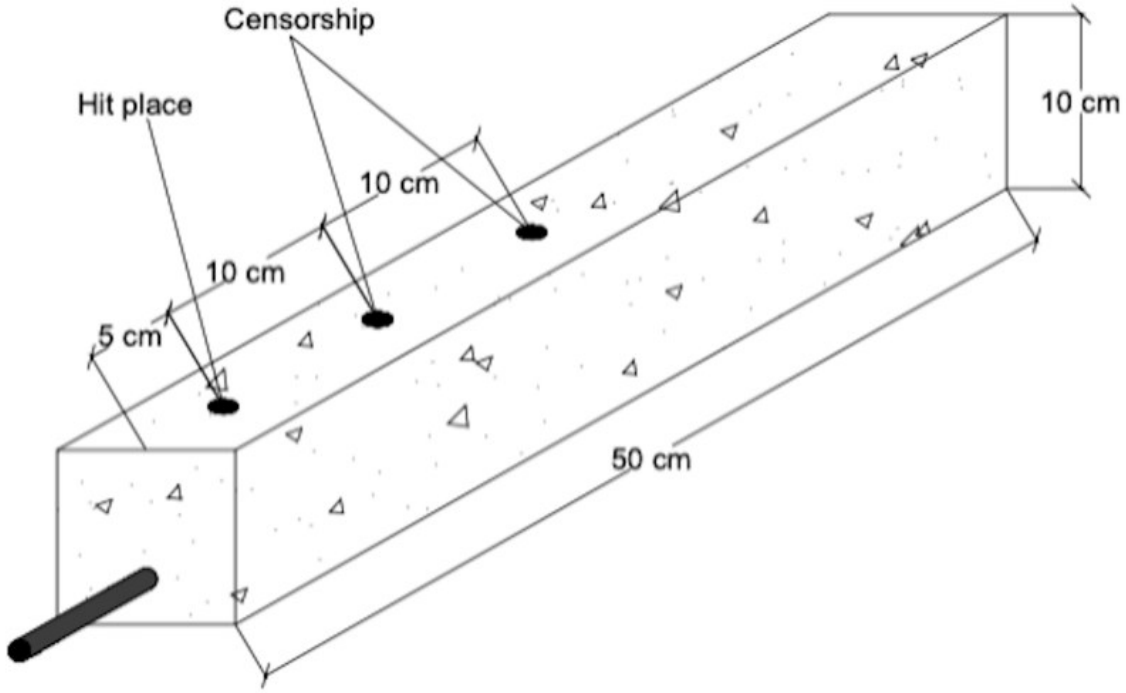


Figure 4. Impact echo test illustration

https://doi.org/10.31436/iiumej.v26i3.3606

2.6. Accelerated Steel Corrosion

In this test, the specimen's corrosion acceleration is required owing to the very long natural corrosion process based on ASTM STP866-EB [10]. The steel reinforcement on concrete is corroded using a DC power supply tool to speed up the corrosion process and adjust the duration and use of the electric current as needed. The reinforcing steel was connected to the positive (+) pole as its anode, and the other reinforcing steel was attached to the negative pole (-) as the cathode [11]. The solution had a salinity of 5% and a corrosion percentage of 20% using a DC power supply with a current of 5 A. The water level is 12.5 cm from the bottom of the container. The salinity of the salt solution was maintained at 5%. Before the concrete was immersed, the concrete reinforcement was wired on one side to connect to the positive pole of the DC power supply tool. The salt solution was then soaked into the concrete and reinforcement parts. The reinforcement used on the negative pole was completely submerged. The corrosion acceleration scheme is illustrated in Figure 5.

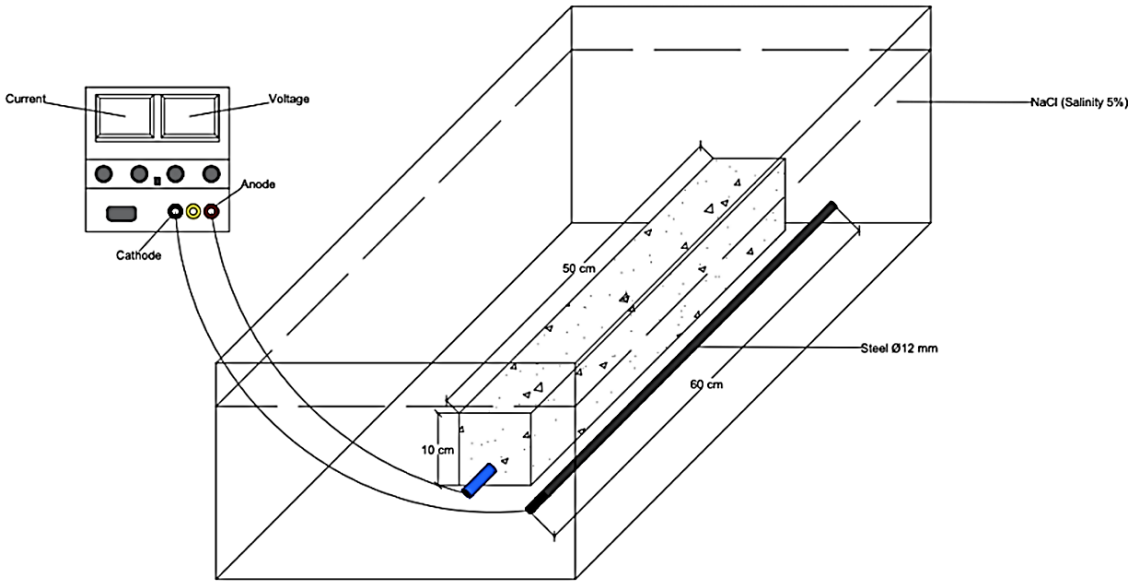


Figure 5. Corrosion acceleration testing scheme

2.7. Concrete Mechanical Testing

Compressive strength testing was performed at the Laboratory of Civil Engineering Structures and Construction Materials, Universitas Muhammadiyah Yogyakarta. This test was performed on 28-day-old cured cylindrical concrete. A Universal Testing Machine is a tool that determines the compressive strength of a test specimen in a concrete mixture using fly ash and silica fume. The compressive strength test scheme is illustrated in Figure 6 (a). The bending strength of concrete was tested using a Universal Testing Machine after the corrosion process was completed. Concrete bending strength testing was performed at the Laboratory of Structure and Construction Materials, Civil Engineering, Universitas Muhammadiyah Yogyakarta. The test was carried out with two loading points in the middle of the beam span and a distance of 5 cm between the pedestal and the edge of the beam. This test aimed to determine the flexural strength of the corroded beams at each percentage of fly ash and silica fume. The flexural strength test scheme is illustrated in Figure 6 (b).

https://doi.org/10.31436/iiumej.v26i3.3606

(a) (b)

Figure 6. (a) Compressive strength testing and (b) Flexural strength testing

RESULTS AND DISCUSSIONS

3.1. Compressive Strength

The results of the compressive strength tests are shown in Figure 7. Fly ash (FA) and silica fume (SF) influence the compressive strength of concrete by enhancing the pozzolanic reaction and improving the microstructure. FA contributes to long-term strength gain through gradual pozzolanic activity. At the same time, SF significantly increases early and later-age strength by filling microvoids and forming additional calcium silicate hydrate (C-S-H), resulting in a denser and stronger concrete matrix. Compressive strength of concrete is planned for Mix Design at 30 MPa. Meanwhile, after testing based on Table 5, the actual compressive strength of concrete was found to be regular concrete without any additives or mixtures, with an average compressive strength of 37.59 MPa. Concrete with 10% FA and 5% SF variations produces an average compressive strength value of 35.10 MPa. The concrete variations of FA 20% and SF 10% show an average compressive strength value of 33.80 MPa. The compressive strengths of the 30% FA and 15% SF concrete yielded an average of 27.44 MPa. Several factors affect compressive strength: the compaction process, which is not optimal, and the material factor, Pozzolan [12]. Adding SF tends to result in higher water absorption, so the workability is low (as shown in the slump value results), making it difficult for fresh concrete to drain correctly into the mold.

https://doi.org/10.31436/iiumej.v26i3.3606

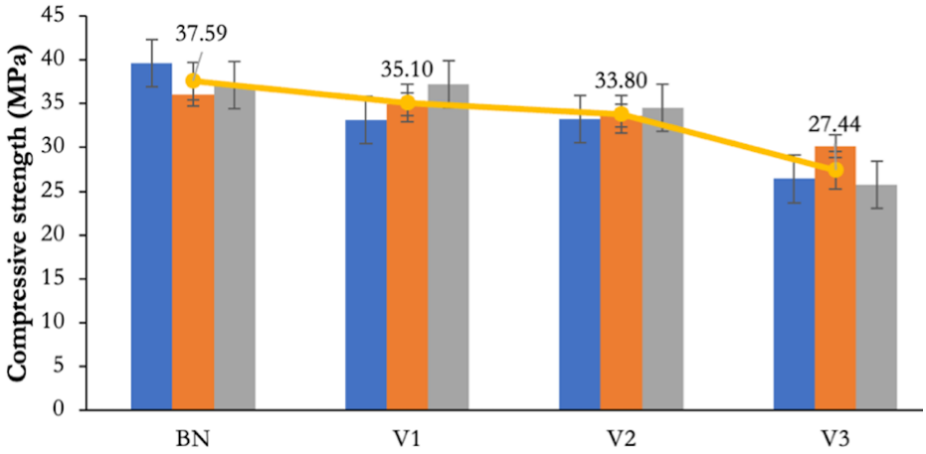


Figure 7. Results of the concrete compressive strength test

The decrease in compressive strength, along with the increase in cement replacement with FA and SF, is due to fly ash having a high silica content, and silica fume causes a decrease in the workability of concrete [13]. The specimen was monitored after testing for compressive strength to observe the crack pattern and determine the behavior of each concrete mixture when subjected to an axial load. According to SNI 1974:2011 [14], there are five types of concrete cracking patterns: cone collapse, cone and split fracture, cone and slide failure, sliding fracture, and vertical-axis parallel fracture. The crack pattern of the specimen is shown in Figs. 8 to 10. FA and SF refine the pore structure and improve the bonding within the matrix, leading to higher tensile strength and reduced crack width or spacing while influencing drying shrinkage and thermal properties, which can alter how and where cracks form compared to normal concrete. In addition, higher amounts of SF typically lead to a denser and stronger matrix, resulting in finer, narrower cracks, while increasing FA content improves long-term strength and reduces thermal cracking. Still, excessive FA may delay early strength gain, potentially leading to wider or more irregular cracking at early ages if not properly cured.

Figure 8 shows the collapse pattern in cylindrical concrete with variations of 10% FA and 5% SF. Figs. 8(a), (b), and (c) show different crack patterns. Figure 8(a) and (b) show the same crack pattern, namely cone and shear, whereas Figure 8(c) shows the shear crack pattern. The failure mechanism in concrete, particularly the development of cone cracking and shearing, is influenced by several factors, including aggregate distribution, interfacial transition zones (ITZ), and localized stress concentrations. Uneven aggregate size can lead to non-uniform stress distribution within the matrix, but this alone does not comprehensively explain the observed failure modes. Instead, cone cracking is typically initiated by point loading or high compressive stress beneath loading points, where tensile stresses radiate outward. Figure 9 shows the collapse pattern in cylindrical concrete with 20% FA and 10% SF. Figure 9(a) shows a sliding crack pattern, Figure 9(b) shows a columnar crack pattern, and Figure 9(c) shows a cone crack pattern. These different cracking patterns were caused by uneven or poorly adhered capping of the concrete surface, resulting in varying tension concentrations in each test piece. Figure 10 shows the collapse pattern of the cylindrical concrete with 30% FA and 15% SF. Figure 10 (a), (b), and (c) show similar crack patterns, namely shear crack patterns. A shear crack pattern occurs in concrete with low compressive strength and is ductile. Shear crack patterns in concrete have significant implications on structural performance, as they indicate potential failure in the load transfer mechanism, particularly in beams and structural members subjected to transverse loads. These cracks can reduce the stiffness, load-bearing capacity, and

https://doi.org/10.31436/iiumej.v26i3.3606

overall ductility of the structure, potentially leading to premature failure if not adequately addressed through adequate design, reinforcement, or material modification.

(a) (b) (c)

Figure 8. Concrete crack pattern with FA 10% and SF 5%

(a) (b) (c)

Figure 9. Concrete cracking pattern FA 20% and SF 10%

https://doi.org/10.31436/iiumej.v26i3.3606

(a) (b) (c)

Figure 10. Concrete cracking pattern FA 30% and SF 15%

3.2. Corrosion Acceleration Results

Corrosion testing was performed after 28 days of curing. The specimens were subjected to accelerated corrosion for 48, 96, and 168 h. The corrosion acceleration process was carried out by immersing the specimen in a NaCl solution with 5% water volume in a Styrofoam box. Meanwhile, a DC power supply was used to provide an electric current to the concrete reinforcement so that the corrosion acceleration process could occur. Graphs of concrete corrosion acceleration for durations of 48, 96, and 168 h are shown in Figure 11. The corrosion acceleration duration (48 h, 96 h, and 168 h) significantly affected the corrosion rates, with longer exposure times leading to higher corrosion levels due to prolonged electrochemical reactions and chloride penetration. As the duration increased, the degree of rust formation, crack development, and mass loss of reinforcement also intensified, indicating a direct correlation between exposure time and corrosion severity. The results of corrosion acceleration for 48 h show an increasing current trend that continues to increase until the end of the canal in some specimens.

In normal concrete, there is a significant increase until the 17th hour. Still, after that, there is a significant decrease, which is likely to occur because of the condition of the reinforcement that can no longer flow the current-voltage effectively to the entire reinforcement. Among the four specimens, the specimens with a mixture of FA and SF produced a lower initial and average flow than normal concrete. The best durability was observed for the V3 mixture, with an average current of 0.248 A. The results of corrosion acceleration at 96 h showed a fluctuating trend but tended to increase from start to finish. In this test, the best resistance behavior was observed in the V3 test specimen, with an average current of 0.612 A. At 168 h, the corrosion acceleration exhibited a fluctuating trend. In this test, the V1 test specimen showed the best resistance, with an average current of 1.160 A. From the overall test, it was observed that a significant current surge occurred for the V2 test specimen at a corrosion acceleration of 168

https://doi.org/10.31436/iiumej.v26i3.3606

hours, which initially showed the lowest initial current compared to the other specimens. Due to its lower current, concrete with FA and SF has a lower corrosion rate than normal concrete. A lower current indicates a weaker corrosion process. Additionally, FA and SF make the concrete denser and less permeable, limiting the penetration of chloride ions and moisture into the steel reinforcement. Their pozzolanic activity also improves the microstructure and alkalinity, enhancing protection against corrosion.

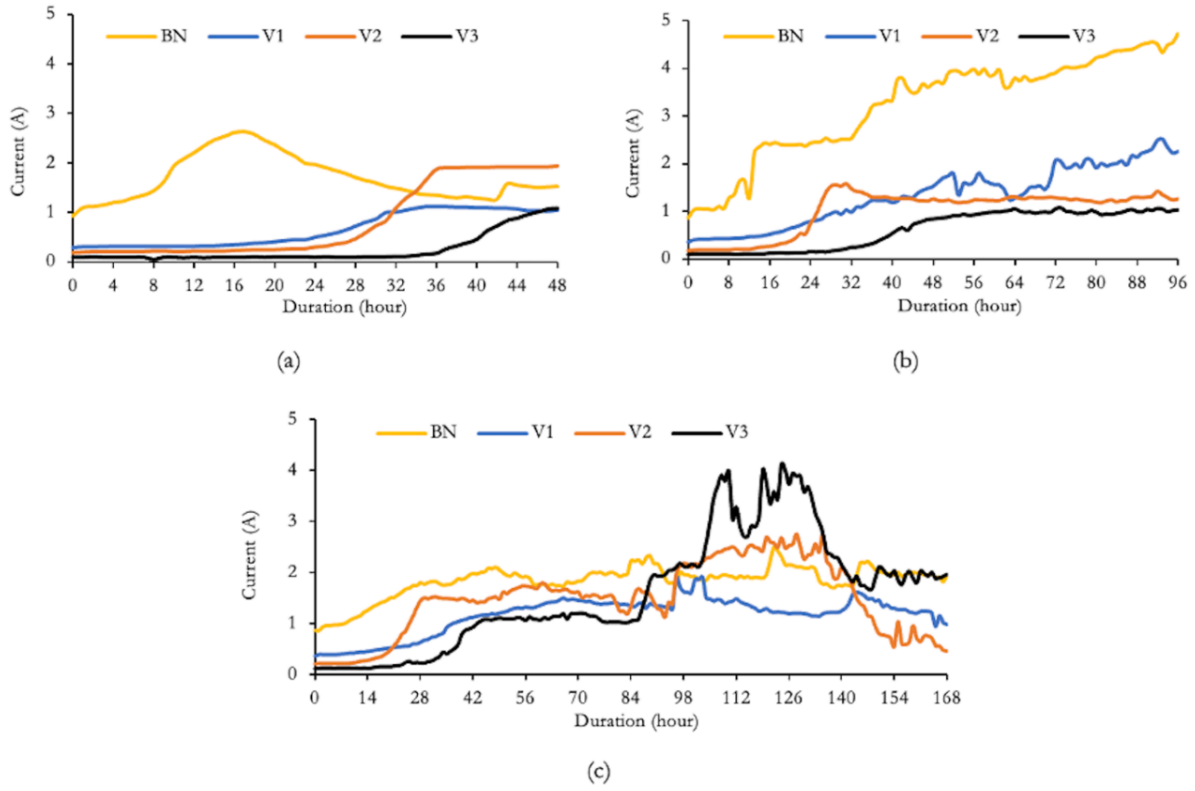


Figure 11. Corrosion acceleration graph (a) 48 h corrosion, (b) 96 h corrosion, and (c) 168 h corrosion

The results of the overall corrosion acceleration show that the initial current in the specimen with pozzolanic material was lower than that of normal concrete [15]. Subsequently, an increase in current was observed every hour owing to continuous corrosion acceleration in all specimens. At 48 h and 96 h corrosion acceleration, V3 has a low average current value. Still, at 168 h corrosion acceleration, V3 has a significant increase in current at 102 h, which is most likely due to the specimen experiencing wide cracks so that the reinforcement is not covered by concrete. After the corrosion acceleration process, cracking in each concrete specimen was observed. Cracks due to this condition can be assessed from crack patterns and width measurements. Concrete crack widths are shown in Figure 12 and Table 4.

https://doi.org/10.31436/iiumej.v26i3.3606

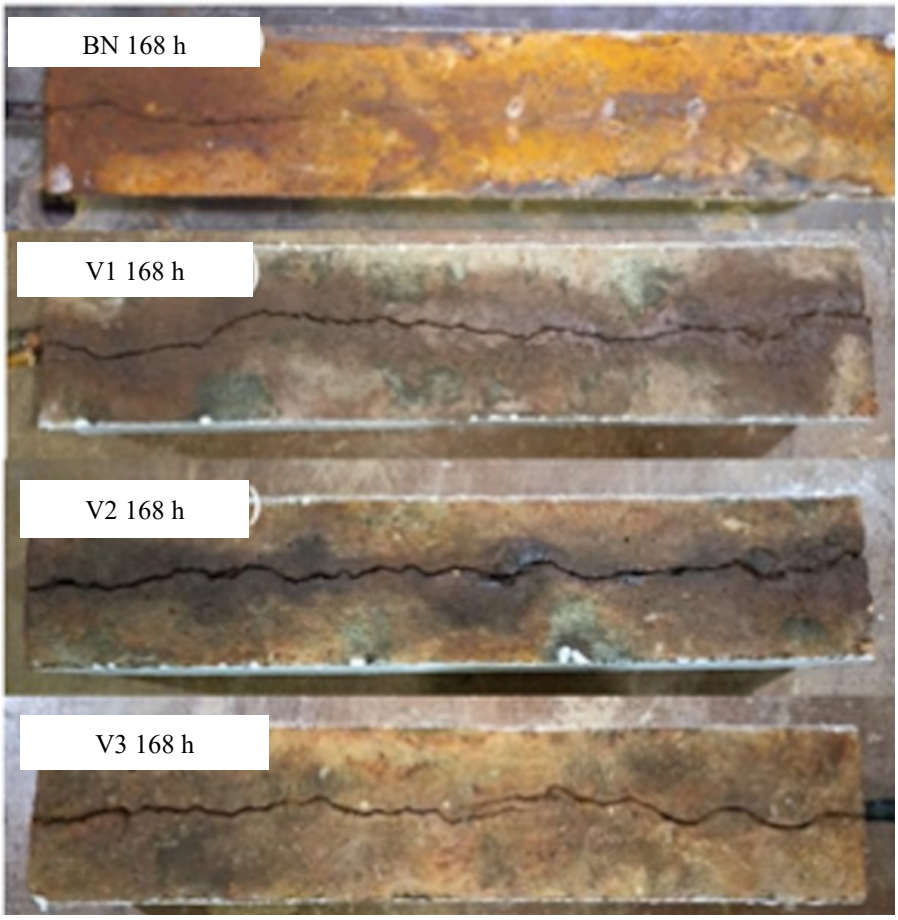


Figure 12. Concrete cracking after the acceleration process

The crack width measurements directly correlated with the corrosion rate in the concrete specimens, as wider cracks generally indicated higher corrosion activity due to the expansive pressure of rust formation. As the corrosion rate increased, the steel reinforcement expanded, causing more significant cracking that further accelerated chloride ingress, creating a feedback loop that worsened the corrosion process. Table 4 shows the width of corrosion cracking on the right and left sides of the concrete specimens. The cracking width is measured using a crack width ruler. The specimens are labeled as a, b, and c to indicate that each concrete variation has three specimens.

Table 4. Corrosion crack width

Test Specimen	Acceleration Duration (hours)	Corrosion crack width (mm)	
		Right side	Left side
BN1	48	0.60	0.30
BN2	168	6.50	0.70
BN3	96	0.75	0.40
V1a	48	0.60	0.60
V1b	96	5.00	0.85
V1c	168	0.85	0.95
V2a	48	0.60	0.60
V2b	96	0.75	0.70
V2c	168	1.40	1.30
V3a	48	0.30	0.20
V3b	96	1.10	1.40
V3c	168	1.40	3.50

https://doi.org/10.31436/iiumej.v26i3.3606

According to the observation results, a crack pattern occurred in all specimens, characterized by cracks in the longitudinal direction. Stiffness was measured on the right and left sides of the concrete. Cracks after 168 hours of corrosion duration had a comparatively larger width. This occurs because the longer the corrosion acceleration process lasts, the more rust products are produced. According to Alhawat et al. [16], the longer the corrosion process lasts, the more rust accumulates on the surface of the steel rod, and the faster the corrosion is owing to cracking, allowing a large number of electrons to move from the cathode to the anode. In addition, the durability of concrete is also an essential factor in the occurrence of cracks. The measurement results indicate that the largest cracks are most frequently observed on the right side, which is primarily characterized by cables. This is possible because the reinforcement closest to the cable has a more positive potential; therefore, electrons occur faster.

After the corrosion acceleration process and bending strength test were completed, concrete destruction was performed to determine the final rebar weight after the corrosion test. After removing the reinforcing steel from the concrete, the reinforcing iron was cleaned of rust due to the corrosion acceleration process. The reinforcing steel was weighed to determine the mass of the remaining rebar due to the corrosion process. Table 5 shows the actual mass of the concrete reinforcement due to the corrosion acceleration process. The mass loss of the reinforcement indicated that adding FA and silica fume SF effectively reduced corrosion, as specimens containing FA and SF showed significantly lower mass loss compared to normal concrete. This demonstrated their ability to enhance concrete durability, minimize chloride penetration, and better protect the embedded steel reinforcement. Table 5 shows that the actual corrosion rate of a normal concrete specimen (BN3) is 34.60%, representing a difference of 24.60% from the planned corrosion rate. Variation 1 (V1b) specimen produced an actual corrosion rate value of 18.33%, with a difference from the planned corrosion of 8.33%. In the variation 2 (V2b) specimen, the actual corrosion value was obtained at 14.17%, with a difference from the planned corrosion of 4.17%. Meanwhile, the variation 3 (V3b) specimen produced an actual corrosion value of 8.75%, different from the planned corrosion value of 1.25%.

Table 5. Percentage of rebar mass loss

Test Specimen	Corrosion Acceleration Duration (hours)	Initial Mass of Reinforcement (grams)	Mass After Corrosion (grams)	Actual Mass Loss (grams)	Actual Corrosion Rate (%)	Percentage of Mass Loss Difference (%)
BN1	48	500	454	46	9.20	4.20
BN2	168	500	287	213	42.60	27.60
BN3	96	500	327	173	34.60	24.60
V1a	48	480	461	19	3.96	1.04
V1b	96	480	392	88	18.33	8.33
V1c	168	480	344	136	28.33	13.33
V2a	48	480	457	23	4.79	0.21
V2b	96	480	412	68	14.17	4.17
V2c	168	480	332	148	30.83	15.83
V3a	48	480	472	8	1.67	3.33
V3b	96	480	438	42	8.75	1.25
V3c	168	480	374	106	22.08	7.08

From these results, it can be observed that the actual corrosion level produces different values for each variation. In addition, it is known that the corrosion rate that occurs in concrete with a mixture of FA and SF is lower than in normal concrete. This is in line with research by Gomez-Luna [17], which shows that the performance of FA and SF can prevent corrosion

https://doi.org/10.31436/iiumej.v26i3.3606

effectively. This can also be seen from the results for normal concrete (BN3), which shows the highest corrosion value. Then, in concrete variation 1 (V1b), the corrosion rate decreased by 16.27% with the addition of variation fly ash and silica fume by 10% and 5%. In variation 2 (V2b), the corrosion rate decreased by 4.17% with the addition of 20% and 10% fly ash and silica fume, respectively. Meanwhile, variation 3 (V3b) reduced the corrosion rate by 12.50%, adding FA and silica fume SF by 30% and 15%, respectively. Corrosion severely impacts the long-term durability of concrete structures by causing steel reinforcement to expand due to rust formation. This expansion leads to cracking, spalling, and a weakened bond between steel and concrete. Over time, the resulting deterioration reduces structural integrity, lowers load-bearing capacity, and shortens the structure's service life, leading to higher maintenance costs and potential safety risks.

3.3. Resistivity Results

A resistivity test was performed after the curing process was completed. The curing method involves soaking the concrete removed from the formwork in a soaking tub for 28 days. The resistivity test was performed at four points on the surface of the specimen. It is important to perform resistivity testing before and after the corrosion acceleration process to evaluate changes in the concrete's ability to resist ionic movement, which directly relates to its corrosion resistance [18]. The test results in resistivity before corrosion show a corrosion category at a very low level on test specimens with FA and SF additives, with a value range of 37-254 kΩ.cm. Normal concrete is in the low and very low corrosion category in the range of 21-37 Ω.cm. Test results show resistivity after corrosion; each specimen has a different duration of corrosion acceleration. After the specimen was corroded, there was a decrease in the resistivity, which decreased with the duration of corrosion acceleration. In specimens with FA and SF, additives were in the moderate to very low range. All normal concrete specimens were included in the moderate to low range. Resistivity is a measure of a material's ability to resist the flow of ions in an electric field. The higher the resistivity value, the better the concrete's ability to prevent corrosion. The resistivity of the test results is shown in Figure 13.

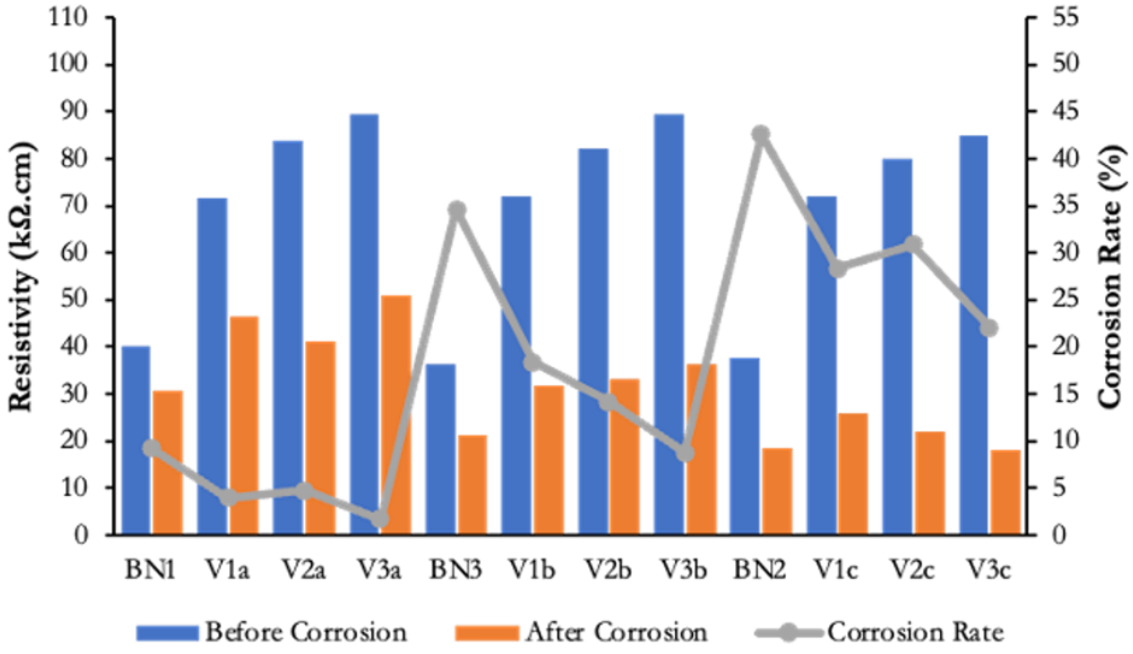


Figure 13. Relationship graph of resistivity and corrosion rate

https://doi.org/10.31436/iiumej.v26i3.3606

The environmental conditions and characteristics of the concrete mixture influence the resistivity of concrete. If the initial resistivity of concrete is high, it will provide a high current resistance, such that the corrosion level in concrete is not too high. In this study, the FA and SF specimens have higher resistivity values than normal concrete. A good corrosion resistance was observed in each FA and SF mixed concrete, and the best performance was observed in the V3 specimens. The addition of FA and SF materials can improve the bonding quality and pore structure of concrete [19]. The pore structure is the most crucial characteristic that controls the transport properties of hydrated cement paste. Therefore, with the addition of fly ash, a pozzolanic reaction can affect water permeability. Silica fume particles are smaller than cement, thus filling the space between the cement grains, which can cause the concrete to become denser and reduce water permeability [20]. The results showed that adding fly ash and silica fume can provide adequate corrosion resistance. In addition, the water-to-cement (w/c) ratio is a key factor in determining the corrosion resistance of concrete. A lower w/c ratio creates a denser, less permeable concrete, limiting the penetration of harmful agents like chloride ions and moisture. In contrast, a higher w/c ratio increases porosity, making the concrete more vulnerable to corrosion by allowing easier access for corrosive elements to the steel reinforcement.

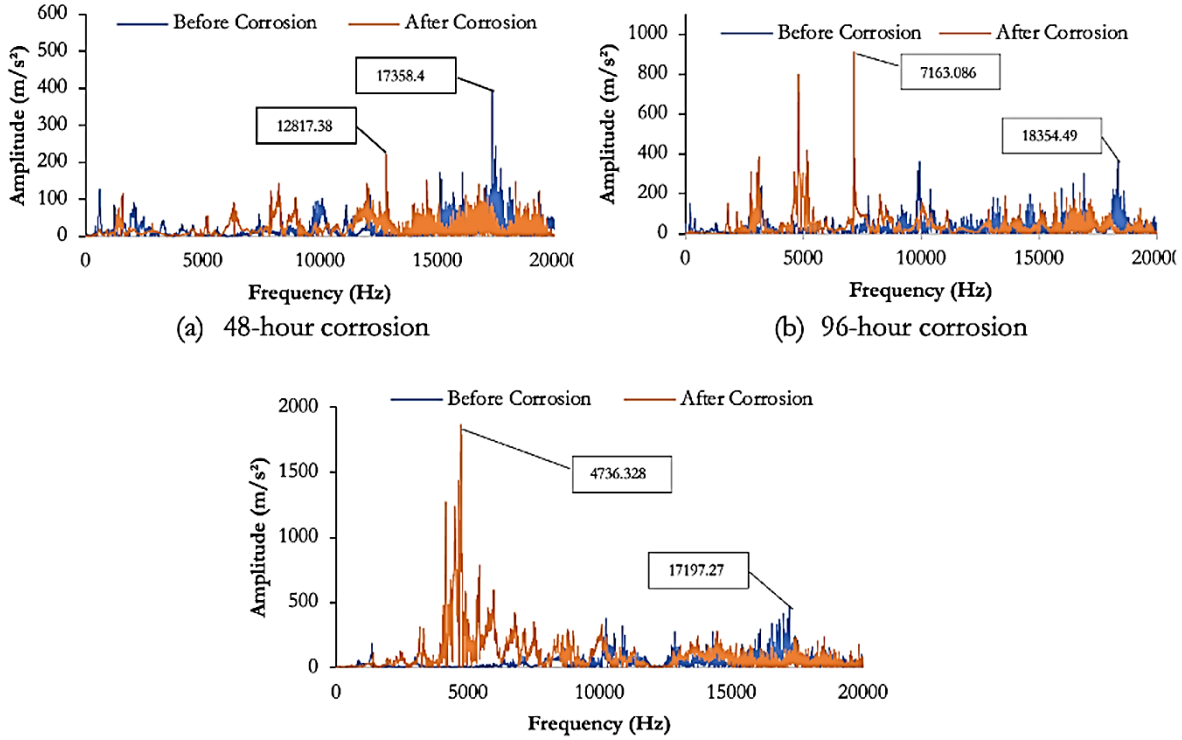
3.4. Impact Echo Results

The impact echo test measures the quality of concrete by evaluating the propagation of stress waves through the material; reflections from internal flaws, voids, or cracks are used to assess concrete integrity, density, and thickness. The frequency is the number of vibrations occurring in one second, a unit of Hertz (Hz). The frequency obtained from the impact-echo test defines the quality of the concrete tested, and the applied frequency results are highly dependent on the distance between the specified sensors; therefore, the author tested four different sensor distances to obtain constant frequency results to make it easier to identify the impact-echo test results. Graphs of impact echo for the durations of 48 h, 96 h, and 168 h, respectively, are shown in Figure 14. From the graph displayed, it can be seen that the peak frequency shift was obtained from the impact-echo test at each corrosion percentage. This decrease in frequency is caused by a reduction in the elastic modulus of concrete owing to corrosion. The results for the peak frequency values are presented in Figure 15.

The test results showed a dominant frequency variation in each specimen before and after corrosion at different sensor positions. In this study, the specimens with a mixture of FA and SF exhibited a relatively higher frequency. The peak frequency in normal concrete ranges from 2000 Hz to 10000 Hz, whereas in the FA and SF specimens, it ranges from 5000 Hz to 18000 Hz. The highest average frequency was observed in V1, with an average value of 18726.9 Hz at a sensor distance of 10 cm. Meanwhile, normal concrete exhibited the highest average frequency value. Meanwhile, in normal concrete, the highest average frequency value of 10214.8 Hz was shown in the BN1 specimen. An extensive evaluation of corroded concrete over 48, 96, and 168 h showed higher FA and SF mixture frequencies. At 168 h, the V2c specimen produced a frequency value of 7712.4 Hz, while BN2 produced a frequency value of 2621.3 Hz. This indicates that a more severe corrosion rate in reinforced concrete can decrease the frequency value. In addition, the amplitude is usually higher in the internal damage area, which is characterized by more significant damage or cracks [16]. Both specimens showed more area noise after being exposed to corrosion than before the measurements. According to Faris et al. [21], corrosion can increase reflected wave amplitude by spreading rust into concrete cracks. At the same time, the presence of chloride and moisture simultaneously significantly attenuates the amplitude and frequency. The combination of FA and SF mixtures has a positive effect in reducing water absorption and chloride permeability, as well as increasing the

https://doi.org/10.31436/iiumej.v26i3.3606

hardness of concrete; the frequency peak is higher than that of normal concrete [2]. The concrete specimens containing fly ash (FA) and silica fume (SF) had fewer internal defects and more uniform wave responses, indicating improved microstructure and overall quality compared to normal concrete.



(a) 48-hour corrosion (b) 96-hour corrosion

(c) 168 hours corrosion

Figure 14. Graphs impact of echo before and after corrosion

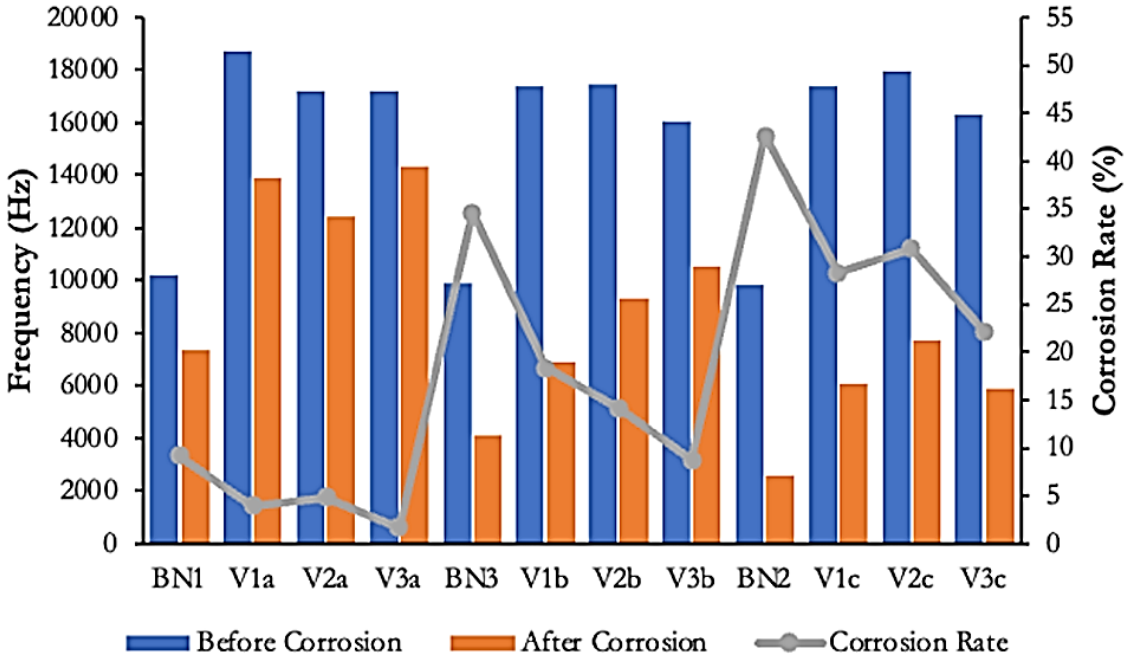


Figure 15. Relationship between frequency value and corrosion rate

https://doi.org/10.31436/iiumej.v26i3.3606

Due to increased density and stiffness from pore refinement and microstructure improvement, concrete with FA and SF mixtures exhibited higher frequencies than normal concrete. This enhancement enabled better wave propagation, resulting in higher resonant frequencies in the impact echo test, which signifies superior material quality and fewer internal defects. In contrast, corroded specimens exhibited a decline in frequency, indicating internal deterioration caused by crack formation, stiffness loss, and increased porosity resulting from corrosion-induced reinforcement expansion. This frequency shift is a reliable indicator of structural damage and degradation within the concrete matrix. The best corrosion resistance was observed for V3. Corrosion significantly affected the frequency values in the impact echo test. The expansion of rust led to crack formation and internal damage, disrupting wave propagation and causing lower, more scattered frequency readings. This decline indicated a loss of concrete stiffness and integrity, highlighting the extent of deterioration caused by corrosion. A greater frequency shift corresponds to higher corrosion levels, as increased internal cracking and stiffness loss further reduce the resonant frequency. This makes frequency shift a reliable indicator for assessing corrosion damage in concrete structures.

3.5. Flexural Strength Results

The bending strength of the reinforced concrete was determined by bending strength testing. The bending strength test was carried out using the three-point bending test method, in which concrete is concentrated on two pedestals with 1-point loading. The bending strength testing was performed after the concrete had been corroded. The results of the flexural strength tests with a mixture of FA and SF variations are shown in Figure 16.

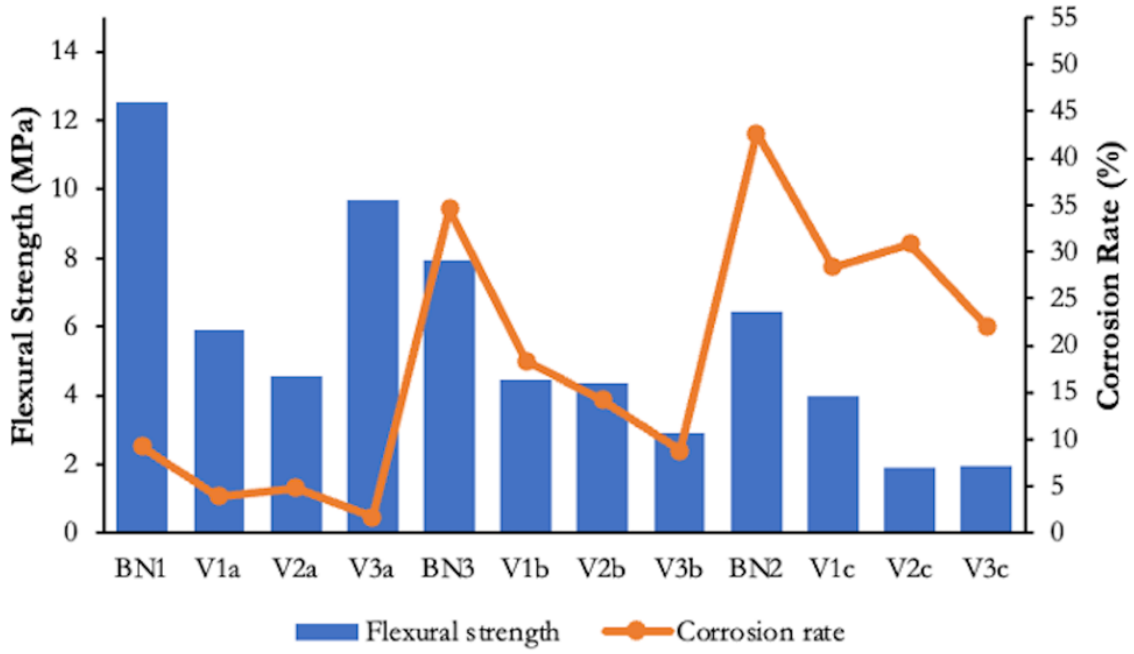


Figure 16. Flexural strength value

Figure 17 shows that the increased corrosion resulted in a decrease in the strength of all specimens. The highest bending strength was observed on the V3 specimen, with a bending strength of 9.69 MPa at a corrosion rate of 1.67%. Meanwhile, at the highest position, BN obtained a bending strength of 12.53 MPa and a corrosion rate of 9.2%. In the 96-hour corroded specimen, the BN specimen at a corrosion rate of 34.6% produced a bending strength of 7.96MPa. Meanwhile, the V2 specimen, at the lowest corrosion rate of 4.79%, had a lower bending strength than BN, with a value of 4.56 MPa. In this case, the corrosion rate did not

https://doi.org/10.31436/iiumej.v26i3.3606

significantly influence the strong bending performance. The best performance was observed for the V3a specimen. The 168-hour corroded specimen showed a different pattern in the V3c specimen, with the lowest bending strength value of 1.93 MPa at a corrosion rate of 22.08%. In contrast, V1c, with a higher corrosion rate of 28.33%, produced a higher flexural strength value of 3.99 MPa.

From the results of this test, it can be concluded that the level of corrosion in each specimen is not the only factor affecting the bending strength of concrete. This is because each specimen exposed to corrosion over time will show a decrease in strength owing to corrosion cracking. The addition of FA and SF can inhibit the corrosion rate, which is characterized by a lower corrosion rate than normal concrete because both decrease the porosity of concrete. However, excessive cement replacement with FA and SF leads to a more significant decrease in strength and water absorption. According to Luo et al. [22], the use of FA is not conducive to increasing the strength and durability of concrete. Still, adding SF can increase the compressive and flexural strengths of concrete. However, as the use of FA and SF increases, the strength of concrete decreases [23]. In this test, the crack pattern of each corroded beam specimen was observed. The appearance of this crack was due to the direction of the crack being perpendicular to the axis of the beam. The bending crack occurred in the middle of the span.

3.6. Crack Patterns

The initial crack in the BN specimen was a bending crack, as shown in Figure 17. Based on the observation of the crack pattern, an initial crack appeared at the bottom of the beam near the load. The cracks widened as the load increased, but no other cracks were observed in the beam. The beam collapsed after deformation with a deflection of 2.604 mm in beam BN1 (48 h corrosion), in beam BN3 2.154 mm (96 h corrosion), and 1.426 mm in beam BN2 (168 h corrosion).

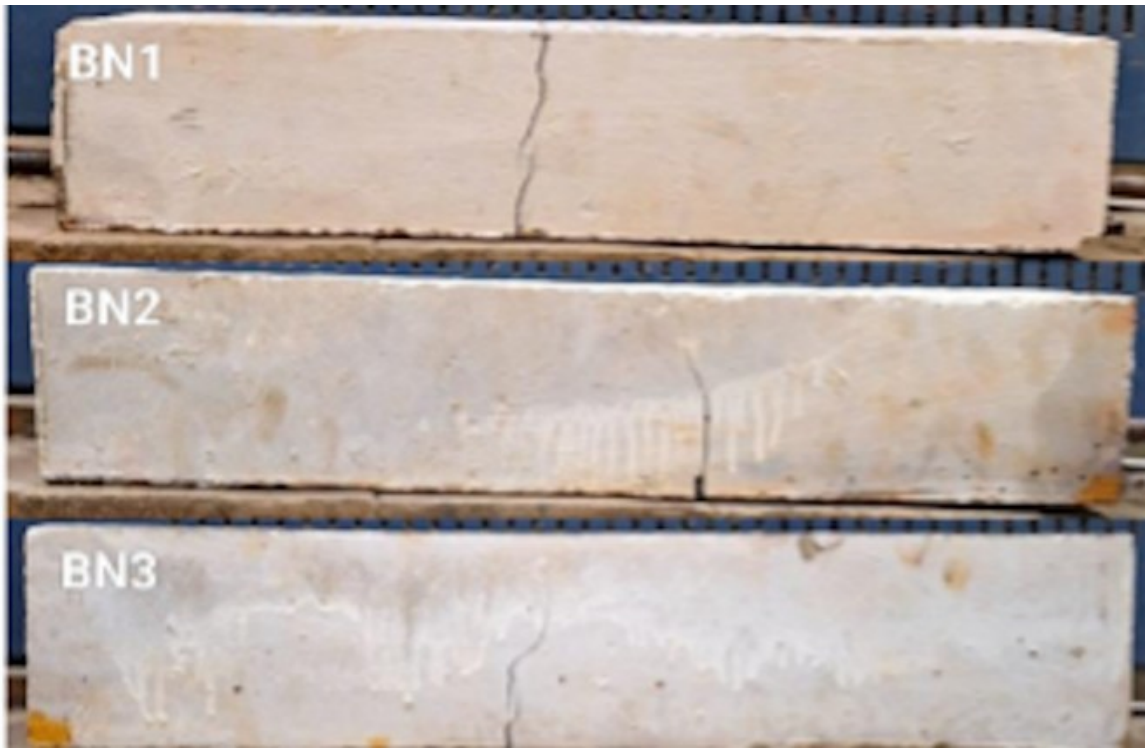


Figure 17. BN beam crack pattern

https://doi.org/10.31436/iiumej.v26i3.3606

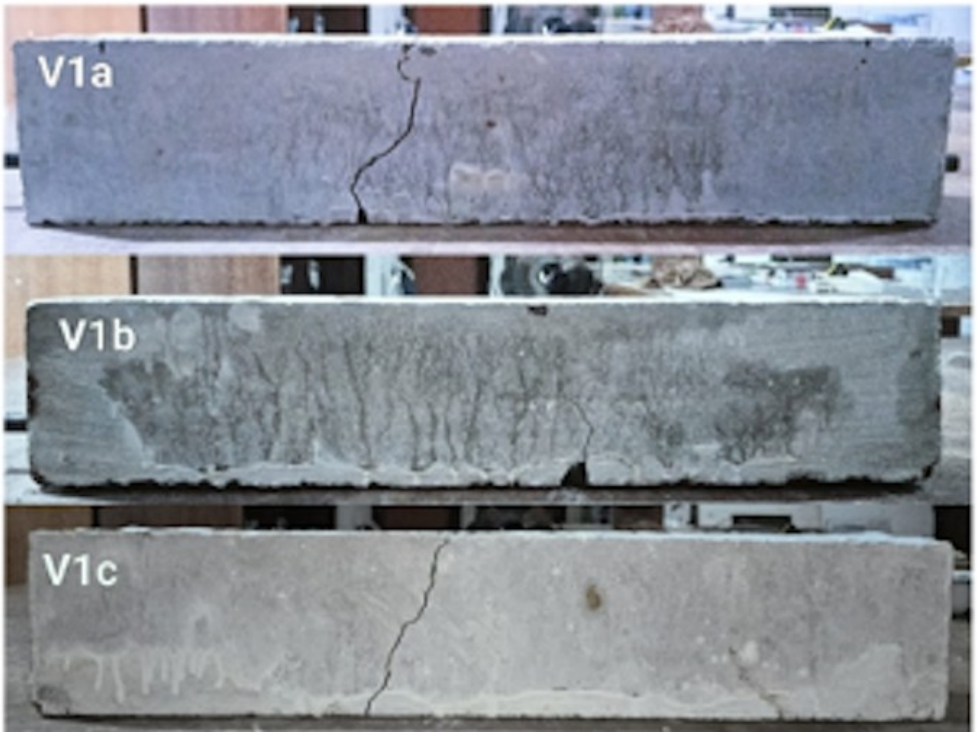


Figure 18. V1 beam crack pattern

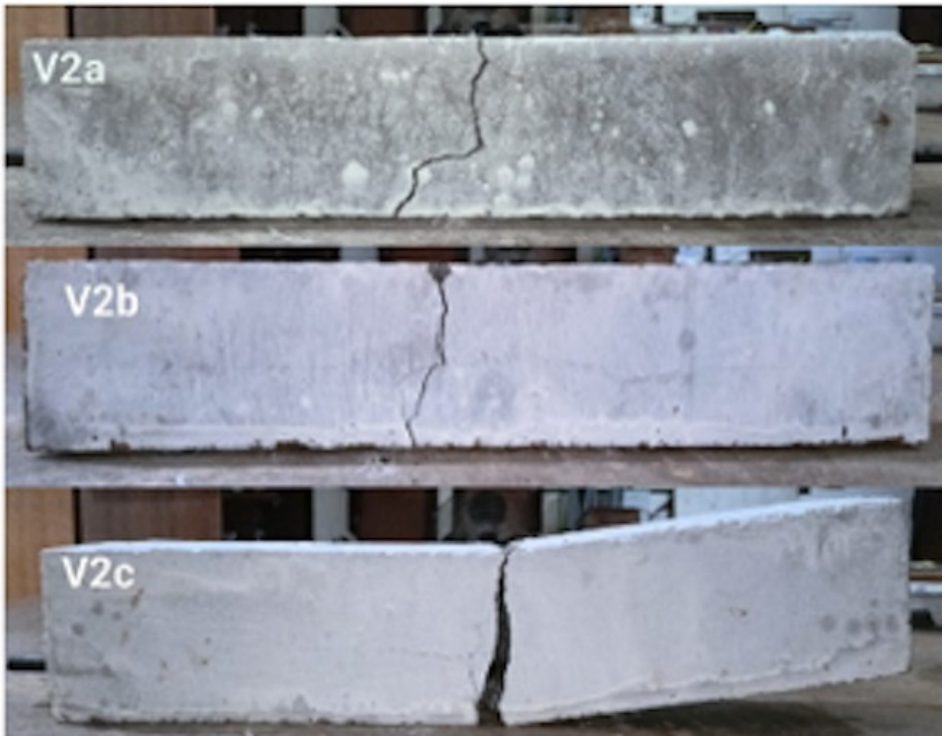


Figure 19. V2 beam crack pattern

The initial crack in the V1 specimen was flexural, as shown in Figure 18. Based on the observation of the crack pattern, an initial crack appeared in the area under the beam close to the load. As the load increased, the cracks widened, and no other cracks were observed along the beams. The beam collapsed after deformation, with a deflection of 3.544 mm on the V1a

https://doi.org/10.31436/iiumej.v26i3.3606

beam (48 h of corrosion), 1.636 mm on the V1b beam (96 h of corrosion), and 1.264 mm on the V1c beam (168 h of corrosion).

The initial crack in the V2 specimen was flexural, as shown in Figure 19. Based on the observation of the crack pattern, an initial crack appeared in the area under the beam close to the load. As the load increased, the cracks widened, and no other cracks were observed along the beams. The beam collapsed after deformation with a deflection of 4.944 mm on the V2a beam (48 h corrosion), on the V2b beam of 2.596 mm (96 h corrosion), and 15.720 mm on the V2c beam (168 h of corrosion).

Figure 20. V3 beam crack pattern

The initial crack in the V1 specimen was flexural, as shown in Figure 20. Based on the observation of the crack pattern, an initial crack appeared in the area under the beam close to the load. As the load increased, the cracks widened, and no other cracks were observed along the beams. The beam collapsed after deformation with a deflection of 5.342 mm on the V3a beam (48 h corrosion), 0.910 mm on the V3b beam (96 h corrosion), and 1.610 mm on the V1c beam (168 h corrosion).

3.7. Microstructure Test Results

SEM testing was performed on specimens corroded for 168 h, and with the best flexural strength, namely, the V1c specimens. The test sample was placed close to the rebar and cut to a maximum size of 1 × 1 × 1 cm. Two test specimens were used for comparison in the analysis: normal concrete samples and concrete containing a mixture of FA and SF. The test results are in the form of photos that show the elements contained in the concrete and can also be seen as the presence of pores or cracks. SEM images are shown in Figure 21.

https://doi.org/10.31436/iiumej.v26i3.3606

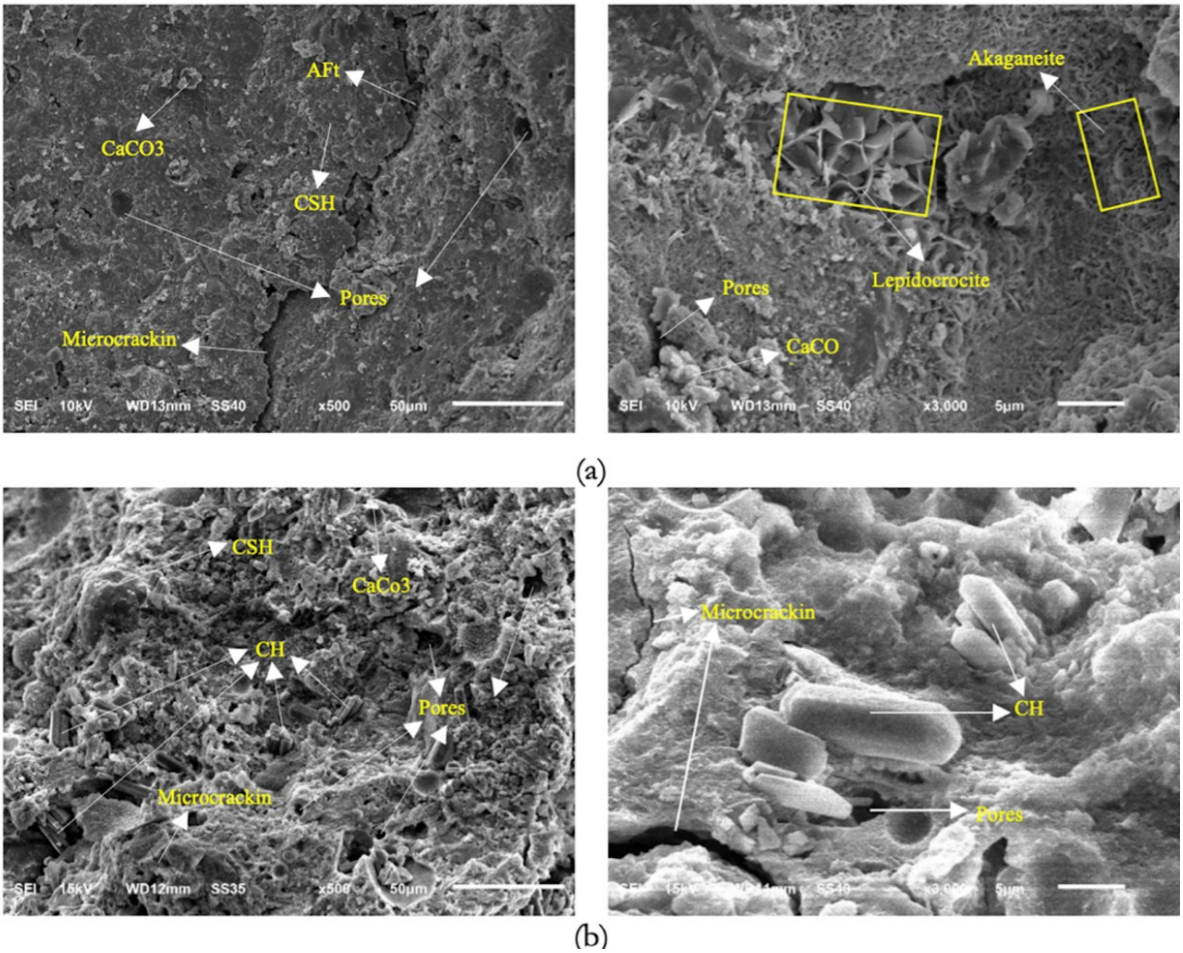


Figure 21. Concrete microstructure (a) BN2 and (b) V1c

In this test, heterogeneous and intermittent distribution of CSH tissue was observed in the V1c specimens, whereas CSH tissues were more homogeneous in the BN2 specimens. This shows that V1c has a weaker structure than that of BN2. The heterogeneous and intermittent distribution of CSH networks on V1c directly leads to a decreased strength compared with BN2, which has a more homogeneous CSH network [24, 25]. The XRD results also showed the identified crystal phase peaks, as shown in Table 7 and Figure 22. XRD testing showed a degree of BN2 of 66.4% for the V1c sample and 72.3%. The degree of crystallinity affects the concrete structure, which is stronger in terms of the concrete's compressive and bending strengths. This suggests that the V1c sample has a composition that is mostly more amorphous than BN2, which can result in lower strength. According to Liu et al. [26], insufficient hydration levels can affect the amount of CSH and the poor crystallinity of the material, thereby reducing the strength of the concrete.

Table 6. Percentage of crystallinity and amorphous Sample

Sample	Crystallinity (5%)	Amorphous (%)
BN2	66.4	33.6
V1c	72.3	27.7

https://doi.org/10.31436/iiumej.v26i3.3606

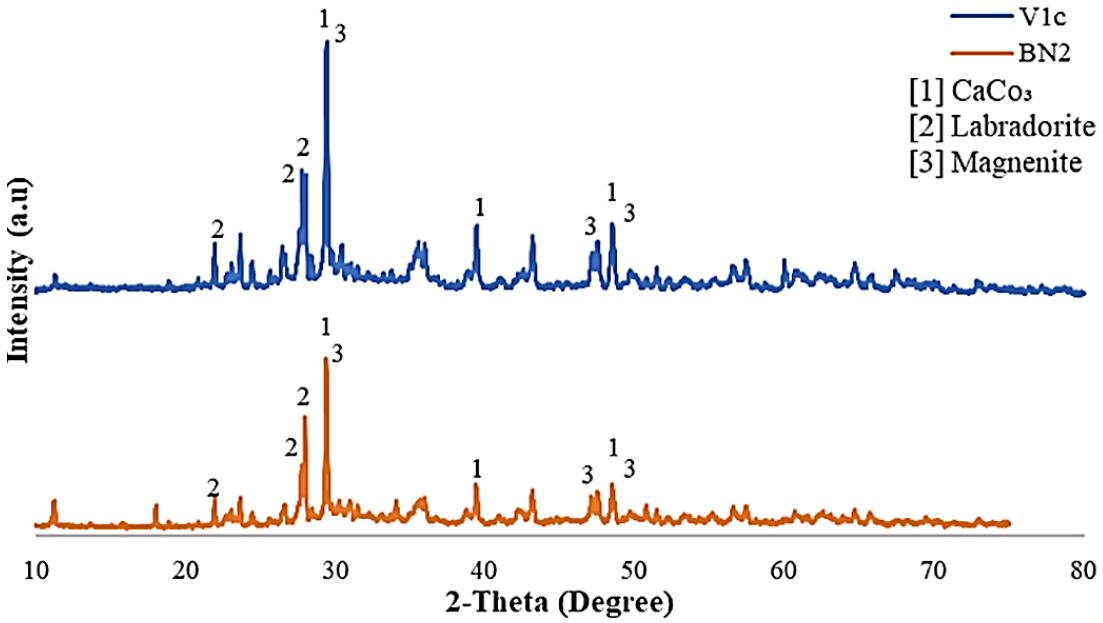


Figure 22. XRD analysis graph

1. CONCLUSIONS

Based on the experimental results and data analysis, the corrosion evaluation of fly ash (FA) and silica fume (SF) concrete using non-destructive testing (NDT) methods reveals several key findings. The compressive strength of concrete generally declined as the percentage of cement replacement with FA and SF increased, with SF in particular contributing to higher water absorption, lower workability, and reduced strength. Over time, all specimens exhibited a decrease in flexural strength due to corrosion-induced cracking, which was further influenced by mechanical degradation resulting from weaker grain adhesion and delayed pozzolanic reactions compared to conventional cement hydration. Nonetheless, the inclusion of FA and SF effectively inhibited the corrosion rate, as confirmed by resistivity and impact echo analyses. The FA and SF concrete samples demonstrated higher corrosion resistance than normal concrete, with pre-corrosion resistivity values ranging from 37–254 kΩ·cm (very low corrosion risk), compared to 21–37 kΩ·cm for normal concrete (low corrosion risk). Additionally, peak frequency values from impact echo testing were notably higher in FA and SF mixes (16,000–18,000 Hz) than in normal concrete (8,000–12,000 Hz), indicating denser internal structures. Both resistivity and impact echo values declined with prolonged exposure to corrosion acceleration. Among all mixes, the V3 specimen, containing 30% FA and 15% SF, exhibited the best overall corrosion performance, balancing mechanical integrity and corrosion resistance most effectively.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Education, Culture, Research, and Technology (Indonesia), Grant number 107/E5/PG.02.00.PL/2024, and by Universitas Muhammadiyah Yogyakarta.

REFERENCES

[1] Nindhita KW, Zaki A, Zeyad AM. (2024) Effect of Bacillus Subtilis Bacteria on the mechanical properties of corroded self-healing concrete. Frattura ed Integrita Strutturale, 18(68):140-58.

[2] Zaki A, Fikri MA, Wibisono CA, Rosyidi SAP. (2023) Evaluating Pre-Corrosion and Post-Corrosion of Oil Palm Shell Concrete with Non-Destructive Testing. Key Engineering Materials, 942:137-62.

[3] Zaki A, Chai HK, Aggelis DG, Alver N. (2015) Non-destructive evaluation for corrosion monitoring in concrete: A review and capability of acoustic emission technique. Sensors (Switzerland), 15(8):19069-101.

[4] Saraswathy V, Muralidharan S, Thangavel K, Srinivasan S. (2003) Influence of activated fly ash on corrosion-resistance and strength of concrete. Cement and Concrete Composites, 25(7):673-80.

[5] Baltazar-Zamora MA, D MB, Santiago-Hurtado G, Mendoza-Rangel JM, Gaona-Tiburcio C, Bastidas JM, et al. (2019) Effect of Silica Fume and Fly Ash Admixtures on the Corrosion Behavior of AISI 304 Embedded in Concrete Exposed in 3.5% NaCl Solution. Materials (Basel), 12(23):4007.

[6] Adamu M, Mohammed BS, Liew MS. (2018) Mechanical properties and performance of high volume fly ash roller compacted concrete containing crumb rubber and nano silica. Construction and Building Materials, 171:521-38.

[7] ACI. (1991) Standard Practice for Selecting Proportions for Normal, Heavyweight, and Mass Concrete. ACI Committe, 1-38.

[8] AASHTO. (2014) AASHTO TP 95-14 : Standard Test Method for Surface Resistivity of Concrete's Ability to Resist Chloride Ion Penetration. Washington, D.C: AASHTO.

[9] ASTM. (2022) C1383 : Standard Test Method for Measuring the P-Wave Speed and the Thickness of Concrete Plates Using the Impact-Echo Method. West Conshohocken, PA: ASTM International, 1-11.

[10] ASTM. (1985) ASTM STP866-EB : Laboratory Corrosion Tests and Standards. ASTM International, 1-640.

[11] Hong S, Zheng F, Shi G, Li J, Luo X, Xing F, et al. (2020) Determination of impressed current efficiency during accelerated corrosion of reinforcement. Cement and Concrete Composites, 108:103536.

[12] Singh M, Siddique R. (2015) Effect of Low-Calcium Coal Bottom Ash as Fine Aggregate on Microstructure and Properties of Concrete. ACI Materials Journal, 112(5), 693-703.

[13] Williams KC, Balamuralikrishnan R. (2023) Strength prediction and correlation of concrete by partial replacement of fly ash & silica fume. Advances in Concrete Construction,16(6):317-25.

[14] BSN. (2011) SNI 1974:2011: Cara uji kuat tekan beton dengan benda uji silinder (How to test the compressive strength of concrete using cylindrical test specimens). Jakarta: Badan Standarisasi Nasional, 1-20.

[15] Mermerdaş K, İpek S, Anwer AM, Ekmen Ş, Özen M. (2021) Durability performance of fibrous high-performance cementitious composites under sulfuric acid attack. Archives of Civil and Mechanical Engineering, 21(4):147.

[16] Alhawat M, Khan A, Ashour A. (2020) Evaluation of steel corrosion in concrete structures using impact-echo method. Advanced Materials Research, 1158:147-64.

[17] Gómez-Luna G, López-Calvo H, Bremner T, Fajardo-San Miguel G, Castro-Borges P, Montes-García P. (2023) Assessment of corrosion prevention methods for steel reinforcement embedded in concrete exposed to a natural marine environment. Construction and Building Materials, 385:131514.

[18] Zaki A, Rahayu H, Nugraha DA, Rosyidi SAP, Kencanawati NN, Fonna S, editors. (2024) Resistivity Method Evaluation of Corroded OPS-Concrete. E3S Web of Conferences; EDP Sciences.

[19] Sadrmomtazi A, Tahmouresi B, Kohani Khoshkbijari R. (2018) Effect of fly ash and silica fume on transition zone, pore structure and permeability of concrete. Magazine of Concrete Research, 70(10):519-32.

[20] Panjehpour M, Ali AAA, Demirboga R. (2011) A review for characterization of silica fume and its effects on concrete properties. International Journal of Sustainable Construction Engineering and Technology, 2(2).

[21] Faris N, Zayed T, Abdelkader EM, Fares A. (2023) Corrosion assessment using ground penetrating radar in reinforced concrete structures: Influential factors and analysis methods. Automation in Construction, 156:105130.

[22] Luo S, Liang W, Wang H, Wang W, Zou R. (2021) Durability evaluation of concrete with multiadmixtures under salt freeze-thaw cycles based on surface resistivity. Advances in Materials Science and Engineering, 2021:1-18.

[23] Wang J, Che Z, Zhang K, Fan Y, Niu D, Guan X. (2023) Performance of recycled aggregate concrete with supplementary cementitious materials (fly ash, GBFS, silica fume, and metakaolin): Mechanical properties, pore structure, and water absorption. Construction and Building Materials, 368:130455.

[24] Hlobil M, Sotiriadis K, Hlobilova A. (2022) Scaling of strength in hardened cement pastes-Unveiling the role of microstructural defects and the susceptibility of CSH gel to physical/chemical degradation by multiscale modeling. Cement and Concrete Research, 154:106714.

[25] Chen J, Li Z, Jin X. (2017) Morphological effect of C–S–H gel on the elastic behaviour of early-age cement-based materials. Magazine of Concrete Research, 69(12):629-640.

[26] Liu B, Zhang Q, Zhao Y, Huan D, Liu X. (2023) Influence of high altitude on the strength of aerated concrete. RSC advances, 13(17):11762-11770.

https://doi.org/10.31436/iiumej.v26i3.3631

CHARACTERISATION AND STRENGTH ACTIVITY INDEX OF ECO-PROCESSED POZZOLAN

ABRAHAM HENRY BERNARD[1], AHMAD NURFAIDHI RIZALMAN[2,3]*

[1]Faculty of Engineering, Universiti Malaysia Sabah
Jalan UMS, 88400, Kota Kinabalu, Sabah, Malaysia

[2]Nanomaterials Research Centre, Universiti Malaysia Sabah
Jalan UMS, 88400, Kota Kinabalu, Sabah, Malaysia

[3]Nanotechnology Cementitious Materials Research Lab, Universiti Malaysia Sabah
Jalan UMS, 88400, Kota Kinabalu, Sabah, Malaysia

**Corresponding author: ahmadnurfaidhi@ums.edu.my*

(Received: 2 March 2025; Accepted: 30 May 2025; Published online: 9 September 2025)

ABSTRACT: Eco-processed pozzolan (EPP) is a newly emerging pozzolanic material produced from the extraction, refinement, and calcination of spent bleaching earth. Due to its pozzolanic characteristics, EPP is a novel interest in the material sustainability of concrete technology in Malaysia today. However, there is still a lack of studies on how much EPP replacement fulfills the minimum requirements of 75% of the value of the strength activity index (SAI) as specified in the ASTM C618 standard. This study investigated the chemical, physical, mineralogical, and microstructural properties of EPP. The results revealed that EPP fulfilled the criteria as a class N pozzolan as per ASTM C618. According to the XRD result, the mineralogical data suggest that the OPC has a high amount of alite and belite at 65.65% and 10.26%, respectively. In contrast, the EPP showed high amorphous silica at 64.28%. Based on the scanning electron microscopy (SEM) images, the EPP particles had a spherical and porous structure, which increased the amount of water required for the mortar containing EPP. This study also examined EPP's strength activity index (SAI) at 10%, 20%, 30%, 40%, 50%, and 60% as partial cement replacement. The results indicated that the SAI was reduced with high EPP replacement percentages due to lower CaO content. The 28-day SAI value satisfied the minimum requirement of pozzolan as per ASTM C618 for cement replacement by up to 40%. In conclusion, EPP performed well as a cementitious supplementary material, with 10% and 20% being optimal replacements for cement.

ABSTRAK: Pozzolan Terproses-Eko (EPP) merupakan bahan pozzolan yang baharu terhasil daripada proses pengekstrakan, penapisan dan pengkalsinan daripada tanah peluntur terpakai. Disebabkan ciri-ciri pozzolaniknya, EPP menjadi tarikan terbaru dalam teknologi bahan konkrit lestari di Malaysia hari ini. Namun, masih terdapat kekurangan kajian mengenai berapa banyak penggantian EPP yang dapat memenuhi keperluan minimum 75% nilai indeks aktiviti kekuatan (SAI) seperti yang dinyatakan di dalam ASTM C618. Kajian ini menyelidik sifat kimia, fizikal, mineralogi, dan mikrostruktur EPP. Dapatan kajian mendapati bahawa EPP memenuhi kriteria sebagai pozzolan kelas N mengikut ASTM C618. Menurut dapatan XRD, data mineralogi menunjukkan bahawa OPC mempunyai kandungan alite dan belite yang tinggi iaitu sebanyak 65.65% dan 10.26%, masing-masing. Manakala, EPP menunujukkan kandungan silika amorfus yang tinggi pada 64.28%. Berdasarkan imej pengimbas mikroskop elektron (SEM), zarah EPP mempunyai struktur sfera dan berliang, di mana ia meningkatkan jumlah air yang diperlukan untuk mortar yang mengandungi EPP. Kajian ini juga mengkaji SAI EPP pada 10, 20, 30, 40, 50, dan 60% sebagai penggantian separa simen. Dapatan kajian menunjukkan bahawa penurunan nilai SAI dengan peratusan

penggantian EPP yang tinggi disebabkan kandungan CaO yang rendah. Nilai SAI pada 28 hari memenuhi syarat minimum pozzolan mengikut ASTM C618 bagi penggantian simen sehingga 40%. Kesimpulannya, EPP berprestasi baik sebagai bahan tambahan simen, dengan 10% dan 20% sebagai penggantian simen yang optimum.

KEYWORDS: *Eco-Processed Pozzolan, Strength Activity Index, Partial Cement Replacement, Water requirement*

1. INTRODUCTION

Concrete is a widely used construction material made of aggregates, water, and cement. However, cement production has a substantial adverse environmental impact, primarily due to high carbon dioxide (CO2) emissions released into the atmosphere during manufacturing [1]. One way to reduce the carbon footprint of construction is by partially replacing cement with pozzolanic materials, which can significantly improve the performance of concrete and reduce its cost [2-4].

Eco-processed pozzolan (EPP), also known as spent bleaching earth ash (SBEA) [5,6], processed SBEA [7], and calcined SBEA [8], is a newly emerging pozzolanic material and can be used as a cement replacement in concrete. EPP originated from bleaching earth, a pozzolanic mineral used in the production of palm oil. It is incorporated into the refining processes to convert crude palm oil into refined, bleached, and deodorised palm oil. Further extraction, refinement, and calcination of the SBE produce EPP [9,10]. The material was manufactured from a refining plant at Lahad Datu, Sabah.

Many studies have reported that EPP has a large silica (SiO2) content ranging from 42.19% to 60.4% [5-11]. This makes EPP a good pozzolanic material that can partially replace cement in concrete production. Abd Rahman et al. (2020) examined the pozzolanic reactivity of EPP using the strength activity index (SAI) test. The results showed that the SAI values of EPP at both curing ages, 7 and 28 days, were more than the minimum requirement of 75% according to the ASTM C618 standard [10]. This indicates that EPP has sufficient reactivity to qualify as a pozzolan.

Recently, there have been many studies on the use of EPP in regular concrete [12-15], geopolymer concrete [16], pavement block [17], and foamed concrete [7,18]. The studies have revealed that the rate of pozzolanic reaction to produce C-S-H gel increased due to high SiO2 content in EPP. This gel fills the voids in concrete, making the concrete denser, thus improving its strength. Furthermore, the inclusion of EPP in concrete also enhanced its resistance to sulphate, carbonation, and chloride attack [12-14].

Although many studies have been conducted on using EPP as a cement replacement, there is still a lack of studies on how much EPP replacement would fulfil the minimum requirement of 75% SAI value. Therefore, this research aims to examine the effect of incorporating 10, 20, 30, 40, 50, and 60% of EPP on the SAI of mortar.

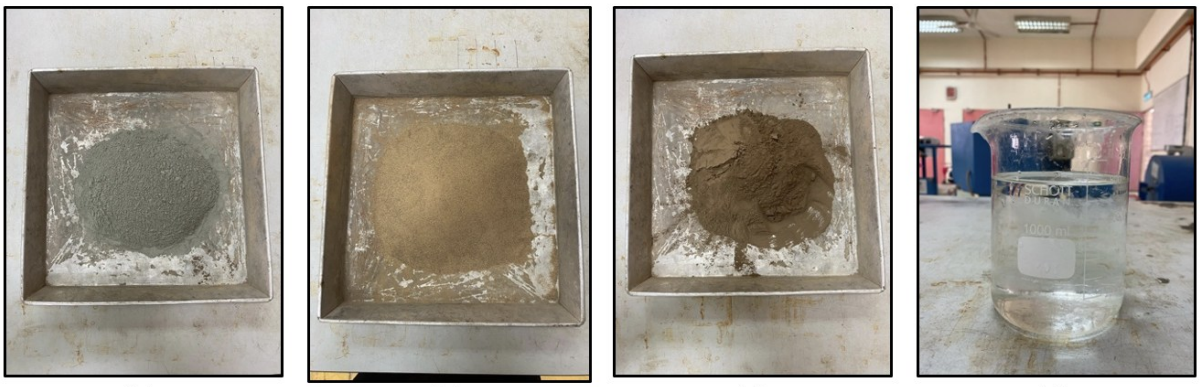
2. METHODOLOGY

2.1. Material

The cement mortars in this study consisted of Ordinary Portland Cement (OPC), EPP, sand, and water (see Fig. 1). The OPC was the Cap Gunung brand from Cement Industries (Sabah) Sdn. Bhd., which complied with MS EN 197-1 CEM I 52, 5N [19]. Meanwhile, the EPP was collected from Eco Oils Sdn. Bhd. is based in Lahad Datu, Sabah.

https://doi.org/10.31436/iiumej.v26i3.3631

The fine aggregate was local river sand sourced from Kota Belud, Sabah. The sand was dried in an oven at 105± 5ºC to remove moisture for 24 hours. Potable water, which complies with ASTM C1602 [20], was used to mix and cure the mortars.



(a) (b) (c) (d)

Figure 1. The samples of (a) OPC, (b) river sand, (c) EPP, and (d) water.

2.2. Mixture Proportions of Mortar Samples

Table 1 shows the mixture proportion of mortars prepared according to ASTM C109 [21]. There were seven (7) specimens, including the control specimen and specimens containing EPP at 10%, 20%, 30%, 40%, 50%, and 60%. The water-to-cement (w/c) ratio used in this study was 0.485 for the control sample.

Table 1. Summary of specimen designation and mixture proportions

Specimen	Cement (g)	Sand (g)	EPP (g)
Control	1000	2750	0
E10	900	2750	100
E20	800	2750	200
E30	700	2750	300
E40	600	2750	400
E50	500	2750	500
E60	400	2750	600

2.3 Testing Procedures

2.3.1. Material Characterization

This study examined OPC and EPP's chemical, mineralogical, microstructural, and physical properties. The material's chemical composition, mineralogical composition, and microstructural properties were determined using X-ray fluorescence (XRF) and X-ray diffraction (XRD), respectively. The specific gravity of the materials was determined following ASTM C188 [22]. Meanwhile, the particle size analyser (PSA) was used to evaluate the particle sizes at d10, d50, and d90. All the characterisation tests for the materials were sent to the reference laboratories for testing, which are more specialised, as tabulated in Table 2.

https://doi.org/10.31436/iiumej.v26i3.3631

Table 2. Summary of tests and laboratories

No	Properties	Tests	Laboratories
1	Chemical composition	XRF	Slag Cement Sdn. Bhd.
2	Mineralogical composition	XRD	Slag Cement Sdn. Bhd.
3	Microstructural properties	SEM	Biotechnology Research Institute (BRI), UMS
4	Specific Gravity	ASTM C188	Geospec Sdn. Bhd.
5	Particle size distribution	PSA	Geospec Sdn. Bhd.

2.3.2. Mortar Sample Preparation and Tests

The mixing procedure was conducted using a mechanical mixer, following the guidelines outlined in ASTM C305 [23]. After mixing, the fresh mortar was poured into a 50 x 50 x 50 mm mould and allowed to cure for 24 hours. Subsequently, the specimens were unmoulded and stored in water for 7 and 28 days. This was done to ensure the specimen attained uniform saturation, minimised leaching, and matured uniformly.

The flowability of the fresh mixtures was measured according to ASTM C1437 [24]. For mortar containing EPP, the water was adjusted to obtain a flow ± 5 of the value of the control mortar, as specified in ASTM C311 [25]. The water required to acquire the flow as per the standard was recorded. Then, the compressive strength of the specimens was measured according to ASTM C109 [21]. Then, the strength activity index (SAI) of EPP at 7 and 28 days was determined according to ASTM C618 [26]. Fig. 2 portrays the instrumentation employed in the study.

(a) (b)

Figure 2. The (a) flowability and (b) compressive strength used to test the properties of the specimens.

3. RESULTS AND DISCUSSION

3.1. Chemical Properties

Table 3 tabulates the chemical oxides of OPC and EPP. EPP's calcium oxide (CaO) content was significantly lower than OPC's, which aligns with past studies [5-11]. The results also reveal that the EPP had higher SiO_2 content than the OPC, with 51.78%, which falls within the range of previous studies [5-11]. Meanwhile, the sum of SiO_2, Alumina (Al_2O_3), and Iron Oxide (Fe_2O_3) was 73.06%. On the other hand, EPP had a lower sulphur trioxide (SO3) content of 0.11%. The results also reveal that the EPP's loss of ignition (LOI) was slightly higher than that of OPC at 3.26%.

From the chemical oxides results, the EPP is classified as Class N pozzolan as per ASTM C618 [26], which is consistent with findings by Lim et al. [6] and Rokiah et al. [7]. Other studies found that EPP was classified as Class C and F pozzolan [5, 9,10]. These differences may be due to the repetitive use of bentonite clay and the type of clay used in the degumming process, which can produce different concentrations of spent bleaching, ultimately making differences in the pozzolan classification of EPP [27].

Table 3. Chemical oxides of OPC and EPP

Chemical Compounds (%)	OPC	EPP	Class N
CaO	62.97	8.47	
SiO_2	19.65	51.78	
Al_2O_3	4.81	10.84	
Fe_2O_3	3.08	10.44	
$SiO_2 + Al_2O_3 + Fe_2O_3$	27.54	73.06	≥ 70.0
MgO	1.75	4.89	
SO_3	2.42	0.11	≤ 4.0
K_2O	0.64	1.4	
Na_2O	0.1	0.2	
P_2O_5	0.13	3.79	
TiO_2	0.31	1.52	
Mn_2O_3	0.19	0.24	
LOI	3.18	3.26	≤ 10.0

3.2. Mineralogical Properties

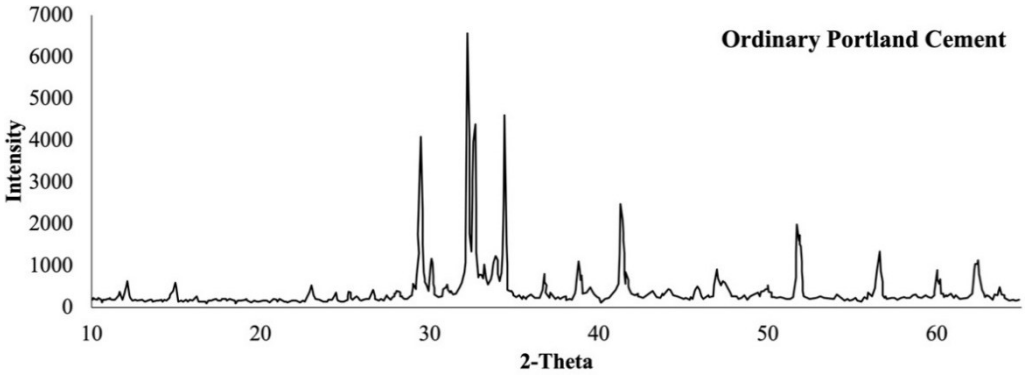
The OPC and EPP XRD analysis is tabulated in Table 4 and Fig. 3, respectively. The OPC comprised a high amount of Alite and Belite, with 65.65% and 10.26%, respectively. These compounds are responsible for developing the early and later strength of concrete. There were also traces of Ferrite (11.33%), Aluminate (3.87%), Calcite (3.51%), Hemi-hydrate (1.81%), and Portlandite (1.16%). Less than 1% of compounds include free lime (fCaO), Periclase, Gypsum, Anhydrite, Langbeinite, Aphthitalite, Arcanite, Lime, Quartz, and Dolomite.

Meanwhile, EPP comprised high pozzolan amorphous silica with 64.28%, indicating pozzolanic reactivity [28]. The second-highest compound in EPP was Quartz, with 15.38%, which may contribute to pozzolanic reactivity [29]. Wafa Mechti found that the pozzolanic reactivity of crystallised quartz with fineness 20um at 10 % Portland cement replacement improved the cement paste compactness by increasing compressive strength and hydration [29]. Less than 10% compounds include Labradorite, Hematite, Magnetite, Calcite, Dolomite, Diopside, Olivine, Hornblende, Cristobalite, and Tridymite.

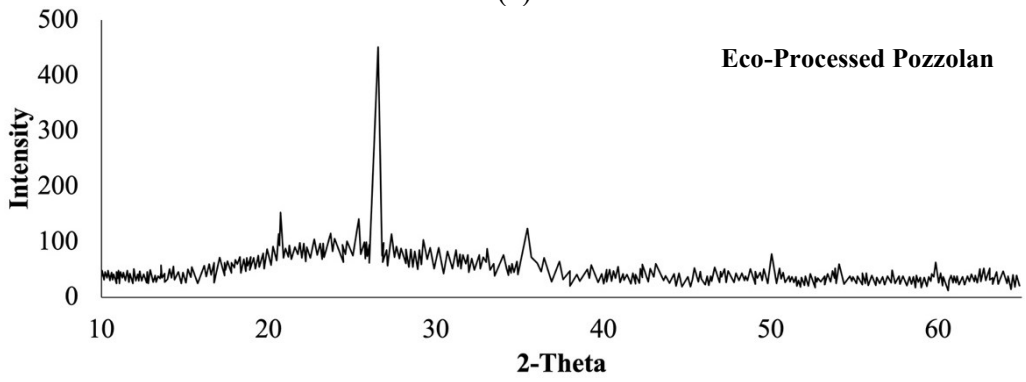
https://doi.org/10.31436/iiumej.v26i3.3631

Table 4: Mineralogical properties of OPC and EPP

Compounds	OPC (%)	EPP (%)
Alite	65.65	-
Belite	10.26	-
Ferrite	11.33	-
Aluminate	3.87	-
Hemi-hydrate	1.81	-
Portlandite	1.16	-
fCaO	0.91	-
Periclase	0.76	-
Gypsum	0.41	-
Anhydrite	0.20	-
Langbeinite	0.18	-
Aphthitalite	0.10	-
Arcanite	0.07	-
Lime	0.03	-
Calcite	3.51	3.24
Quartz	0.43	15.38
Dolomite	0.22	2.27
Pozzolan amorphous Silica	-	64.28
Labradorite	-	6.07
Hematite	-	4.93
Magnetite	-	3.29
Diopside	-	0.08
Olivine	-	0.40
Hornblende	-	0.03
Cristobalite	-	0.02
Tridymite	-	0.01



(a)

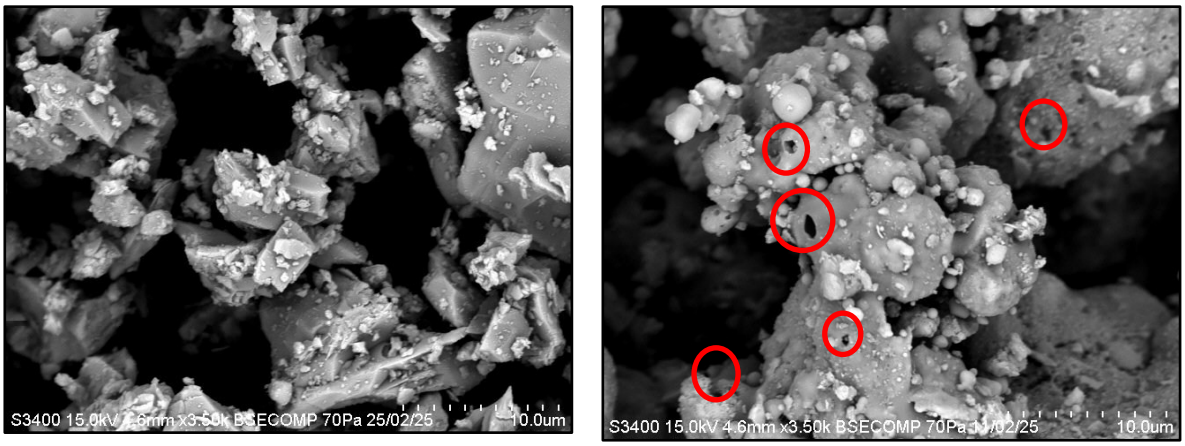


(b)

Figure 3. XRD graph analysis of (a) OPC and (b) EPP

3.3. Microstructural Properties

Fig. 4 presents the SEM images of OPC and EPP, illustrating their respective morphologies. The image reveals that the OPC particles had an angular and irregular structure, consistent with past studies [5,7]. This means that they do not have a uniform and consistent shape. The angular microstructure of the OPC may increase the surface area, which promotes bonding with constituent materials, thus resulting in better strength development [30]. Meanwhile, EPP mainly consists of spherical and porous structures (as indicated in red circles in Fig. 4(b)), with a minor presence of irregular particle shapes, which also aligns with previous studies [5-7,9,10]. The porous microstructure of EPP may inhibit the strength development of the mortar by allowing water to occupy the empty spaces; hence, as it evaporates, it leaves voids that decrease the mortar's strength [31].



(a) (b)

Figure 4. SEM images of (a) OPC and (b) EPP

3.4. Particle Size Distribution

Table 5 represents the physical properties of OPC and EPP. Based on the mean particle size in the table, EPP generally had a larger particle size than OPC. This observation is consistent with Lim et al. [6], who indicated that the EPP was typically coarser than OPC particles. The retained 45-micron sieve showed that EPP retained a smaller percentage than OPC. Coarser particles of EPP in mortar can reduce its compacting density as larger particles of EPP leave larger voids between particles, hindering the close packing of mortar. Lin et al. (2024) found that the increment of EPP replacement on cement reduced the ultrasonic pulse velocity test result, indicating a significant void in the matrix [9].

Based on the results, the specific gravity of OPC and EPP was 3.12 and 1.77, respectively. This indicates that EPP was lighter than OPC, which may be due to its porous structure, as mentioned in Section 3.3.

Table 5. Physical properties of OPC and EPP

Properties	Amount retained 45 μm sieve (%)	Mean particle size			Specific gravity
		d10 (µm)	d50 (µm)	d90 (µm)	
OPC	0.88	0.070	0.162	0.261	3.12
EPP	0.94	0.230	0.354	0.493	1.77

3.5. Flow Spread and Water Requirement of Specimen

The specimen's flow spread and water requirement are tabulated and illustrated in Table 6 and Fig. 5, respectively. The amount of water was adjusted for the mortar with EPP to obtain ±5% of the control mortar as specified in ASTM C311. Based on the result, the water requirement increased with the EPP replacement level. This indicates that high EPP content caused more water to be absorbed by the particles due to their porous texture, as stated in Section 3.3. Moreover, EPP also had a lower specific gravity, which may contribute to the increased demand for water. This finding is consistent with Kho [12], who stated that the finer size and higher porosity of EPP require a higher amount of water, thus causing a slump reduction of concrete.

Table 6. Flow and water requirement of the specimen

Specimen	Flow Spread (mm)	Water Requirement (%)
Control	223	100
E10	223	102
E20	223	106
E30	225	107
E40	223	111
E50	225	114
E60	223	125

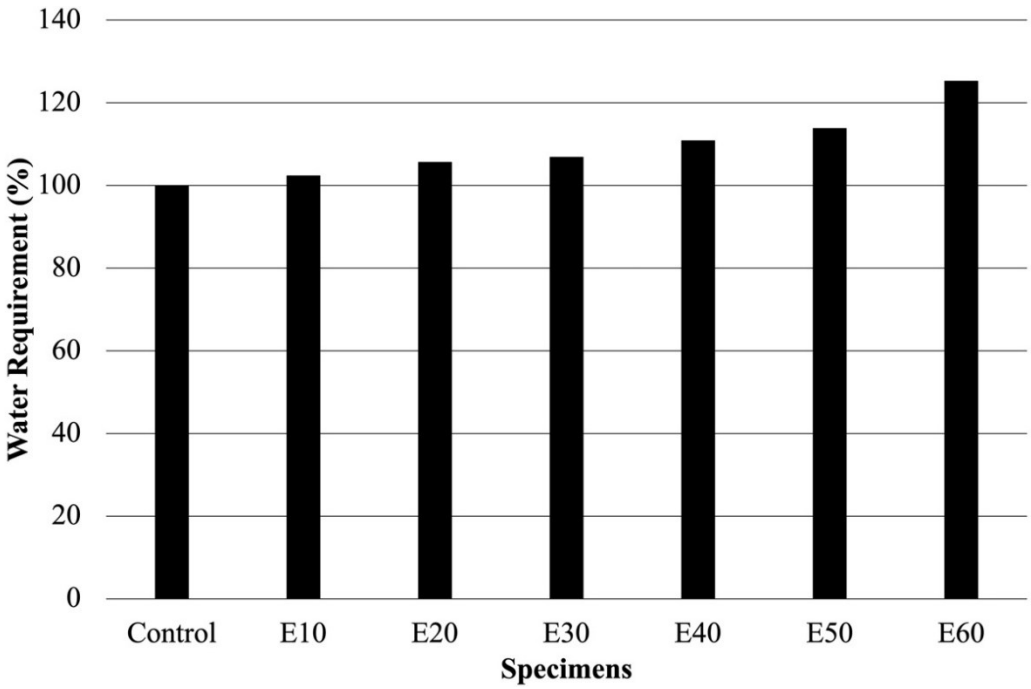


Figure 5. Water requirement of specimens

3.6. Compressive Strength and Strength Activity Index

Table 7 presents the compressive strength of mortar at 7 and 28 days. At 7 days, E10 exhibited the highest compressive strength of 19.2 MPa, which was 13.6% higher than that of the control specimen. Then, there is a decline in strength from E20 to E60, with strength lower than the control by a 3.6% to 75.7% reduction. At 28 days, the compressive strength of E10 and E20 was 14.3% and 4.9% higher than that of the control specimen. However, E30 to E60 had lower compressive strength than the control, declining from 2.7% to 67.6%.

The SAI of the specimens at 7 and 28 days are illustrated in Fig. 5. At 7 days, the SAI values of E10, E20, and E30 were 113.6%, 96.4%, and 78.% respectively, which were more than the minimum requirement of 75% SAI value as stipulated in ASTM C618 [26] (as shown in blue line in Fig. 6). Meanwhile, E40, E50, and E60 did not achieve the minimum requirement. This is because the amount of CaO becomes lower as the amount of EPP increases, which leads to low formation of calcium hydroxide (Ca(OH)2) [5,32,33].

At 28 days, the SAI values of E10, E20, E30, and E40 achieved more than the minimum requirement as per ASTM C618 [26], with 114.3%, 104.9%, 97.3%, and 84.6%, respectively. However, the remaining specimens did not achieve the minimum requirement. According to Abd Rahman et al. [5], the SAI value increment at 7 days and 28 days was due to higher SiO2 content, which enhances the pozzolanic reaction with Ca(OH)2 that occurs at a later stage of strength development [7]. This led to more CSH gel production, thus resulting in high compressive strength [5-7]. Among all the specimens, E10 and E20 had higher SAI values than the control specimens, indicating that 10% and 20% EPP are the optimum levels for cement replacement.

Table 7. Compressive strength and SAI

Specimen	Compressive Strength		SAI (%)	
	7 Days	28 Days	7 Days	28 Days
Control	16.9	18.2	100	100
E10	19.2	20.8	113.6	114.3
E20	16.3	19.1	96.4	104.9
E30	13.2	17.7	78.1	97.3
E40	10.2	15.4	60.4	84.6
E50	7.9	11.2	46.7	61.5
E60	4.1	5.9	24.2	32.4

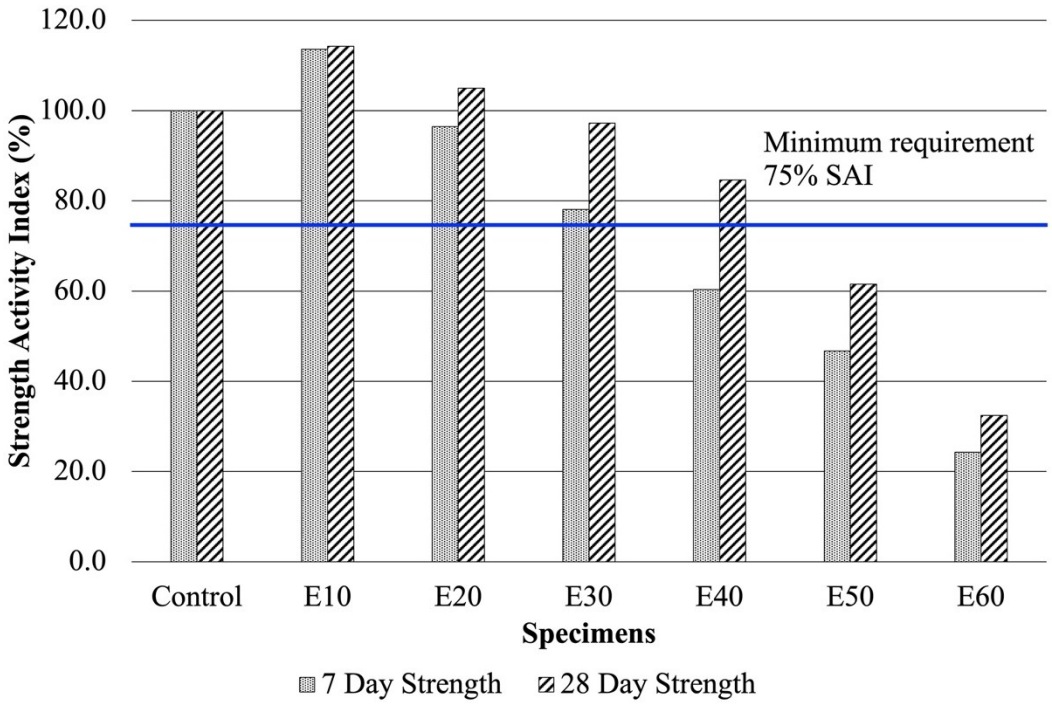


Figure 6. Strength Activity Index (SAI) at 7 and 28 days

In summary, the data obtained from this study may be influenced by environmental variation or experimental error, which could introduce inconsistencies in the conditions under which the samples were prepared, cured, or tested. Factors such as fluctuations in temperature and humidity, variations in mixing techniques, or inconsistencies in the quality of raw materials, especially the eco-processed pozzolan mineralogical and chemical properties, could lead to differences in water demand, water/ binder ratio, or strength development. Additionally, human error during the weighing, mixing, or testing process may have contributed to deviations in the data. These uncontrolled variables can impact the reliability and accuracy of the results, making it difficult to draw definitive conclusions about the actual performance of the pozzolan-blended concrete.

4. CONCLUSION

EPP fulfilled the criteria as a Class N pozzolan as per ASTM C618 and thus can be used as a replacement for cement. The results revealed that the sum of SiO_2, Al_2O_3, and Fe_2O_3 was 73.06 ($\geq$ 70%), the amount of SO_3 was 0.11% ($\leq$ 4%), and the amount of LOI was 3.26% ($\leq$ 10%). The high porosity of EPP particles increased water absorption, which could negatively impact the strength of the mortar.

It was also found that the SAI was reduced with high EPP replacement percentages due to lower CaO content. Meanwhile, the SAI values satisfied the minimum requirement of pozzolan as per ASTM C618 for EPP at cement replacement of 10, 20, 30, and 40%. However, the strength of the mortar was reduced below that of the control specimen when the EPP replacement was up to 30%. Lastly, EPP performed well as a cementitious supplementary material, with 10% and 20% as an optimal cement replacement.

The results have shown that EPP can be used as a partial cement replacement in concrete. This can reduce the reliance on traditional cement in concrete production, which has a higher carbon footprint. Meanwhile, it can also reduce raw material costs and energy consumption in cement manufacturing, ultimately decreasing the overall cost of concrete production. Furthermore, the improved performance of concrete with EPP as partial cement replacement can lead to longer-lasting structures and reduced maintenance needs.

ACKNOWLEDGEMENT

This research was financially supported by the Malaysian Ministry of Higher Education (Grant No.: FRGS/1/2024/TK06/UMS/02/1).

REFERENCES

[1] Benhelal E, Zahedi G, Shamsaei E, Bahadori A. (2013). Global strategies and potentials to curb CO2 emissions in cement industry. Journal of Cleaner Production, 51 (2013) 142-161. http://dx.doi.org/10.1016/j.jclepro.2012.10.049

[2] Verma NK. (2013). Influence of partial replacement of cement by industrial wastes on properties of concrete. Recent Trends in Civil Engineering, Lecture Notes in Civil Engineering 77. https://doi.org/10.1007/978-981-15-5195-6_54

[3] Bansal M, Bansal M, Bahrami A, Krishan B, Garg R, Ozkilic YO, Althaqafi E. (2024). Influence of pozzolanic addition on strength and microstructure of metakaolin-based concrete. PLoS ONE 19(4):e0298761. https://doi.org/10.1371/journal.pone.0298761

[4] Wassouf M, Omran J, Kheirbek A. (2025). Effect of natural pozzolana, pozzolanic sand, and basalt on thermal and mechanical properties of green concrete. Building Engineering 2025, 3(1), 1739. https://doi.org/10.59400/be1739

[5] Abd Rahman RF, Asrah H, Rizalman AN, Mirasa AK, Rajak MAA. (2022). Strength activity index and properties of spent bleaching earth ash. International Journal of GEOMATE, Vol. 23, Issue 97, pp. 82-89. https://doi.org/10.21660/2022.97.3168

[6] Lim A, Mirasa AK, Asrah H, Tian X. (2022). Maximising volume of spent bleaching earth ash (SBEA) pozzolan used as cement replacement in mortar through mechanical activation. Jurnal Teknologi, 84:5(2022)105-116. https://doi.org/10.11113/jurnalteknologi.v84.18177|

[7] Rokiah O, Khairunisa M, Youventharan D, Mohd Arif S. (2019). Effect of processed spent bleaching earth content on the compressive strength of foamed concrete. IOP Conf. Series: Earth and Environmental Science 244(2019)012013. doi:10.1088/1755-1315/244/1/012013

[8] Lim CH, Anwar SS. (2024). Assessment of pozzolanic reactivity for calcined spent bleaching earth ash (SBEA) using strength activity index (SAI), frattini test and x-ray diffraction (SRD). MATEC Web of Conferences 397, 03003(2024). https://doi.org/10.1051/matecconf/202439703003

[9] Lin Y, Alengaram UJ, Ibrahim Z, Ibrahim MSI, Alnahhal AM, Srinivas MK. Performance appraisal of high volume of sustainable unprocessed eco-minerals, ground and unground eco-processed pozzolan as cement replacement materials in mortar. (2024). Construction and Building Materials, 411(2024)134344. https://doi.org/10.1016/j.conbuildmat.2023.134344

[10] Abd Rahman RF, Asrah H, Rizalman AN, Mirasa AK, Rajak MAA. (2020). Study of eco-processed pozzolan characterisation as partial replacement of concrete. Journal of Environmental Treatment Techniques, Volume 8, Issue 3, Pages 967-970.

[11] Sutrisno W, Alrasyid H, Wulandari KD, Hakam M. (2021). Experimental investigation on properties of concrete mortar incorporating spent bleaching earth waste as supplementary cementitious material. IOP Conf. Series: Earth and Environmental Science 871(2021)012005. doi:10.1088/1755-1315/871/1/012005

[12] Kho JH. (2021). Incorporation of eco process pozzolan (EPP) as partial cement replacement and superplasticisers in concrete. IOP Conf. Series: Earth and Environmental Science 682(2021) 012014. doi:10.1088/1755-1315/682/1/012014

[13] Yunus E, Asrah H, Rizalman AN. (2019). Compressive strength of eco-processed pozzolan concrete under chloride and sulphate exposure. Journal of Advanced Research in Applied Mechanics, 59 Issue 1(2019)1-9.

[14] Yunus E, Asrah H, Rizalman AN. (2020). Strength and sorptivity of eco-processed pozzolan concrete under chloride and sulphate exposure. International Journal of Advanced Research in Engineering Innovation, Vol. 2, No. 3, 32-43.

[15] Abd Rahman RF, Asrah H, Rizalman AN, Mirasa AK. (2023). Effect of spent bleaching earth ash on sulphate attack resistance of concrete. Materials Today: Proceedings. https://doi.org/10.1016/j.matpr.2023.03.054

[16] Ahmat AM, Alengaram UJ, Shamsudin MF, Alnahhal AM, Ibrahim MSI, Ibrahim S, Rashid RSM. (2023). Assessment of sustainable eco-processed pozzolan (EPP) from palm oil industry as a fly ash replacement in geopolymer concrete. Construction and Building Material, 387(2023)131424. https://doi.org/10.1016/j.conbuildmat.2023.131424

[17] Mohd Kusaimi NF, Hamzah F, Jai J, Md Zaki NA, Ibrahim N. (2020). Compressive strength and water absorption of pavement derived from palm oil eco processed pozzolan (EPP) material as partial cement replacement. AJChe 2020, Vol. 20, No. 2, 205-215. 10.22146/ajche.60230

[18] Sharom NB. (2016). Performance of eco process pozzolan foamed concrete as cement replacement. Bachelor's thesis. Universiti Malaysia Pahang.

[19] Malaysian Standard. (2014). MS EN 197-1:2014 Cement – Part 1: Composition, specifications and conformity criteria for common cement. Department of Standards Malaysia.

[20] ASTM. (2022). ASTM C1602/C1602M-22 – Standard Specification for Mixing Water Used in the Production of Hydraulic Cement Concrete. ASTM International. West Conshohocken.

[21] ASTM. (2023). ASTM C 109/C109M-23 – Standard Test Method for Compressive Strength of Hydraulic Cement Mortars (Using 2-in. or [50 mm] Cube Specimens). ASTM International. West Conshohocken.

[22] ASTM. (2023). ASTM C188-23 – Standard Test Method for Density of Hydraulic Cement ASTM International. West Conshohocken.

[23] ASTM. (2020). ASTM C305-20 - Standard Practice for Mechanical Mixing of Hydraulic Cement Pastes and Mortars of Plastic Consistency. ASTM International. West Conshohocken.

[24] ASTM. (2020). ASTM C1437-20 – Standard Test Method for Flow of Hydraulic Cement Mortar. ASTM International. West Conshohocken.

[25] ASTM. (2021). ASTM C311-11 – Standard Test Methods for Sampling and Testing Fly Ash or Natural Pozzolans for Use in Portland-Cement Concrete. ASTM International. West Conshohocken.

[26] ASTM. (2023). ASTM C618-22 – Standard Specification for Coal Fly Ash and Raw or Calcined Natural Pozzolan for Use in Concrete. ASTM International. West Conshohocken.

[27] Latisya S, Nastiti SI, Muslich. (2024). Reactivation of used bentonite (Spent Bleaching Earth) and its reuse as adsorbent for CPO (Crude Palm Oil) bleaching process. BIO Web of Conferences 99, 02013 (2024). https://doi.org/10.1051/bioconf/20249902013

[28] Elyasigorji F, Tabatabai, H. (2024). Assessment of select direct and indirect pozzolanic reactivity test outcomes with robust regression and ranking analysis. Case Studies in Construction Materials, 21 (2024) e03441. https://doi.org/10.1016/j.cscm.2024.e03441

[29] Mechti W, Mnif T, Samet B, Rouis MJ. (2012). Effects of the secondary minerals on the pozzolanic activity of calcined clay: case of quartz. IJRRAS 12 (1).

[30] Ghasemi Y, Emborg M, Cwirzen A. (2018). Estimation of specific surface area of particles based on size distribution curve. Magazine of Concrete Research, Vol. 70, Issue 10, pp. 533-540.

[31] Nagaratnam BH, Mannan MA, Rahman ME, Mirasa AK, Richardson A, Nabinejad O. (2019). Strength and microstructural characteristics of palm oil fuel ash and fly ash as binary and ternary blends in Self-Compacting concrete. Construction and Building Materials, 202, 103–120. https://doi.org/10.1016/j.conbuildmat.2018.12.139

[32] Tural HG, Ozarisoy B, Derogar S, Ince C. (2024). Investigating the governing factors influencing the pozzolanic activity through a database approach for the development of sustainable cementitious materials. Construction and Building Materials, 411 (2024) 134253. https://doi.org/10.1016/j.conbuildmat.2023.134253

[33] Poudel S, Bhetuwal U, Kharel P, Khatiwada S, Diwakar KC, Dhital S, Lamichhane B, Yadav SK, Suman S. (2025). Waste glass as partial cement replacement in sustainable concrete: mechanical and fresh properties review. Building, 15, 857. https://doi.org/10.3390/buildings15060857

https://doi.org/10.31436/iiumej.v26i3.3730

MACHINE LEARNING AND RSM FOR STRENGTH FORECASTING IN SUSTAINABLE SCGC

SAMEH FUQAHA[1], AHMAD ZAKI[2*], GUNTUR NUGROHO[2]

[1]*Dept. of Civil Engineering, Postgraduate Studies, Universitas Muhammadiyah Yogyakarta, Yogyakarta, Indonesia*
[2]*Dept. of Civil Engineering, Universitas Muhammadiyah Yogyakarta, Yogyakarta, Indonesia*

**Corresponding author: ahmad.zaki@umy.ac.id*

(Received: 17 May 2025; Accepted: 3 July 2025; Published online: 9 September 2025)

ABSTRACT: This research focuses on the predictive modeling of flexural (Ff) and splitting tensile (Ft) strengths in Self-Compacting Geopolymer Concrete (SCGC) to support sustainable mix design optimization. A curated dataset comprising 544 experimental records was utilized to train and evaluate eight supervised machine learning (ML) algorithms. These included Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forests, Gradient Boosting, CN2 Rule Induction, Naïve Bayes, Decision Trees, and Stochastic Gradient Descent. The predictive performance of each model was assessed using multiple statistical metrics, such as RMSE, R^2, and accuracy percentage. Among the models, SVM and KNN achieved the highest precision, with R^2 values of 0.99 and RMSE as low as 0.10 MPa. Additionally, statistical techniques were applied to identify influential input variables, confirming the dominant role of binder constituents in determining tensile-related strength. The models demonstrated strong generalization on unseen data and minimal sensitivity to activator dosage or curing age. These results validate the effectiveness of ML-driven tools for SCGC prediction and offer a scalable framework for integrating data analytics into sustainable concrete design and performance optimization.

ABSTRAK: Kajian ini memfokuskan kepada pemodelan ramalan bagi kekuatan lenturan (Ff) dan tegangan belahan (Ft) dalam Konkrit Geopolimer Pemadat Kendiri (SCGC) bagi menyokong pengoptimuman reka bentuk campuran mampan. Satu set data terpilih yang merangkumi 544 rekod eksperimen telah digunakan bagi melatih dan menilai lapan algoritma pembelajaran mesin (ML) terselia. Algoritma tersebut termasuk Mesin Sokongan Vektor (SVM), K-Nearest Neighbors (KNN), Rawak Forests, Gradient Boosting, CN2 Rule Induction, Naïve Bayes, Pokok Keputusan, dan Stochastic Gradient Descent. Prestasi ramalan setiap model dinilai menggunakan pelbagai metrik statistik seperti RMSE, R^2, dan peratusan ketepatan. Antara model tersebut, SVM dan KNN mencapai ketepatan tertinggi dengan nilai R^2 sebanyak 0.99 dan RMSE serendah 0.10 MPa. Tambahan, teknik statistik turut digunakan bagi mengenal pasti pemboleh ubah input berpengaruh, sekali gus mengesahkan peranan dominan konstituen pengikat dalam menentukan kekuatan berkaitan tegangan. Model yang dibangunkan menunjukkan keupayaan generalisasi yang kukuh terhadap data baharu serta kepekaan minimum terhadap dos pengaktif atau umur pengerasan. Dapatan ini mengesahkan keberkesanan alat berasaskan ML bagi meramal SCGC dan menawarkan kerangka boleh skala bagi mengintegrasikan analitik data ke dalam reka bentuk konkrit mampan serta pengoptimuman prestasi.

KEY WORDS: *Self-Compacting Geopolymer Concrete (SCGC); Flexural Strength Prediction; Splitting Tensile Strength; Machine Learning Models; Response Surface Methodology (RSM)*

https://doi.org/10.31436/iiumej.v26i3.3730

1. INTRODUCTION

Self-Compacting Geopolymer Concrete (SCGC) is an innovative material that merges the high flowability and workability of self-compacting concrete (SCC) with the eco-friendly properties of geopolymer technology. This integration provides multiple benefits, especially by enhancing eco-friendly characteristics and improving structural performance [1]. Unlike conventional Portland cement systems, using alkaline activators, geopolymer binders are formulated by activating aluminosilicate-rich industrial residues, such as fly ash, ground granulated blast furnace slag (GGBS), and metakaolin. This approach significantly curtails the greenhouse gas emissions typically linked to traditional cement manufacturing [2]. The substitution of Portland cement with pozzolanic materials has been shown to lower carbon dioxide emissions and energy consumption throughout the concrete's life cycle [3].

SCGC possesses the unique ability to flow and fill formworks solely under the influence of gravity, removing the necessity for external vibration. This property proves highly advantageous in casting intricate geometries and heavily congested reinforcement areas. Beyond improving ease of placement and construction efficiency, SCGC exhibits remarkable performance in terms of strength, durability, and resistance to elevated temperatures, often matching or exceeding the qualities observed in traditional self-compacting concrete [4]. Furthermore, its excellent chemical stability and minimal shrinkage render it highly appropriate for demanding applications, including coastal infrastructure, industrial facilities, and thermally stressed environments [5].

Despite the promising performance of SCGC, optimizing its mix design to achieve desired fresh and hardened properties remains complex due to the nonlinear interaction among multiple influencing variables [6]. Therefore, modern machine learning (ML) techniques have been increasingly utilized to develop predictive models capable of estimating key mechanical properties. Previous studies have successfully implemented ML models such as Random Forest, XGBoost, Multi-Expression Programming, Genetic Programming, and Artificial Neural Networks to forecast the strength of SCC and geopolymer concretes [7–10].

In recent advancements related to alkali-activated concrete systems, numerous studies have explored the integration of ML and response surface methodology (RSM) for predictive modeling and process optimization. Dong et al. [11] introduced a novel framework utilizing extended support vector regression (X-SVR) to estimate chloride-induced corrosion in alkali-activated slag (AAS) concrete. By combining multi-ionic transport simulations with hydrate-based models of chloride binding, the study offered a robust tool for evaluating corrosion risk with enhanced computational efficiency. Complementing this, Iqbal et al. [12] investigated the rheological behavior of alkali-activated materials (AAMs) incorporating hydrated lime. Various ML algorithms were deployed to predict key properties such as plastic viscosity and yield stress, including stochastic gradient descent, K-nearest neighbors (KNN), and support vector machines (SVM). The KNN model demonstrated the highest predictive accuracy, and RSM was further applied to analyze the influence of mix constituents, revealing that hydrated lime increased viscosity. At the same time, fly ash contributed to reducing the yield stress.

Building on these methodologies, Parhi et al. [13] developed a hybrid ML model that merged three base algorithms with a meta-learning layer, surpassing conventional approaches in terms of performance. In this case, RSM served as an optimization tool to fine-tune variables such as curing temperature, alkaline activator concentration (Na_2SiO_3/NaOH ratio), and superplasticizer dosage, achieving compressive strengths within the range of 39–43 MPa. While these approaches have proven effective, much of the existing research focuses predominantly on compressive strength, often neglecting other essential mechanical properties

[14]. Moreover, a limited comparative evaluation of different ML techniques for diverse performance metrics is crucial for selecting the most suitable model for a given application. Jiang et al [15] applied ensemble learning to predict the mechanical behavior of SCGC containing calcium carbide waste and rice husk ash, highlighting the improved generalization ability of ensemble models over standalone algorithms.

Addressing existing research gaps, this work develops and analyzes combined machine learning approaches to forecast the flexural and splitting tensile strengths of SCGC. A diverse range of ML algorithms is investigated, including ensemble techniques such as boosting and bagging, SVM, KNN, and rule-based classifiers, to identify models that most effectively capture the complex tensile behavior of SCGC. The models are trained on an experimentally validated dataset incorporating mix designs with sustainable binders such as fly ash and ground granulated blast furnace slag (GGBS), ensuring relevance to real-world construction practices. Beyond benchmarking predictive performance, the study emphasizes model interpretability by integrating RSM and sensitivity analysis using the Hoffman & Gardner approach. These methods provide both high-fidelity predictions and clear insights into the influence of key input parameters. Ultimately, by developing interpretable and reliable ML tools, this research supports the practical integration of artificial intelligence in concrete technology and contributes to scalable, environmentally responsible engineering solutions.

2. METHODOLOGY

2.1. Data Compilation and Statistical Summary

In this study, a comprehensive dataset was constructed to facilitate the predictive modeling of SCGC's mechanical properties. A total of 544 experimental records were compiled from peer-reviewed journal articles and technical sources, as in Table 1. These entries represent various mixed designs and curing ages, covering diverse proportions of geopolymer constituents and mechanical performance results.

Table 1. Compiled Experimental Records from Prior Literature Sources

References	Data	References	Data
Mazumder et al. [16]	132	Aliabdo et al. [17]	39
Ngernkham et al. [18]	42	Arun et al. [19]	12
Lavanya et al. [20]	51	Deilami et al. [21]	36
Rangan et al. [22]	52	Nishanth et al. [23]	6
Vijaya et al. [24]	32	Alam et al. [25]	108
Tho-In et al. [26]	9	Bheel et al. [27]	85

Each record in the dataset encapsulates a comprehensive set of input and output variables crucial for ML analysis in the context of geopolymer concrete performance prediction. The input variables include the dosage of GGBS, measured in kilograms per cubic meter (kg/m³), which serves as a primary binder component, and FA, also in kg/m³, commonly used to enhance workability and long-term strength. Additionally, the alkaline activators, represented by NaOH and Na_2SiO_3, both quantified in kg/m³, play a key role in initiating the geopolymerization process. The curing age (Age), expressed in days, reflects the time at which strength tests are conducted, capturing the time-dependent evolution of mechanical properties.

The output variables focus on mechanical performance, specifically the Ff and Ft, expressed in megapascals (MPa). Ff provides insights into the material's resistance to bending under load, while Ft indirectly evaluates its capacity to withstand tensile stresses. These output

parameters are vital for understanding structural behavior and for training ML models to predict the mechanical performance of geopolymer concrete based on its compositional and processing variables.

These variables were selected based on their relevance to geopolymerization mechanisms and their influence on the strength development of SCGC [28]. The compiled data represents various curing conditions, mix ratios, and testing procedures, ensuring diversity and representativeness across the parameter space.

2.1.1. Dataset Preparation and Model Structuring

The dataset was randomly divided into two primary subsets: a training set and a validation set, to promote the development of reliable and unbiased ML models. The training portion, comprising 408 data points (representing 75% of the total dataset), was used to fit the models by enabling them to learn the complex relationships between input features and target outputs. The remaining 25% validation set, comprising 136 samples, evaluated how well the trained models could predict outcomes on previously unseen data, testing their generalizability. This random allocation helps reduce the likelihood of overfitting, ensuring the models remain effective in real-world applications. To further reinforce the model's stability and reliability, 10-fold cross-validation was implemented during the training phase. This method offers additional protection against overfitting, particularly valuable when dealing with multiple output variables and inherent non-linear interactions among the predictors [29].

2.1.2. Statistical Analysis and Correlation Mapping

Initial data exploration involved computing key statistical descriptors for all input features and target outputs, including mean, standard deviation, minimum, maximum, and variance. These metrics were evaluated separately for the training set (408 samples) and the testing set (136 samples) to ensure consistent data distribution across both subsets. The summarized results are presented in Tables 2 and 3. This analysis was crucial for assessing variability, detecting potential outliers, and evaluating the overall range of values that could impact model performance. The training set demonstrated sufficient variability across all features, with GGBS and FA showing high standard deviations (156.8 and 156.6 kg/m³, respectively). At the same time, Ff and Ft had standard deviations of 1.6 MPa and 1.1 MPa, respectively. Similar distributions were observed in the testing set, confirming that the data partitioning preserved the underlying patterns and feature dispersion necessary for reliable model training and validation.

Table 2. Statistical Summary of Training Set (n = 408)

Statistic	GGBS (kg/m³)	FA (kg/m³)	NaOH (kg/m³)	Na_2SiO_3 (kg/m³)	Age (days)	Ff (MPa)	Ft (MPa)
Max	525.6	515.5	95.2	133.6	28.0	7.1	5.2
Min	0.0	0.0	52.7	75.0	7.0	0.2	0.1
Avg	247.3	237.9	68.1	101.8	16.0	3.2	2.2
SD	156.8	156.6	10.6	10.7	8.6	1.6	1.1
Var	0.6	0.6	0.2	0.1	0.5	0.5	0.5

https://doi.org/10.31436/iiumej.v26i3.3730

Table 3. Statistical Summary of Testing Set (n = 136)

Statistic	GGBS (kg/m³)	FA (kg/m³)	NaOH (kg/m³)	Na_2SiO_3 (kg/m³)	Age (days)	Ff (MPa)	Ft (MPa)
Max	512.5	510.2	93.6	132.0	28.0	6.9	5.1
Min	0.0	0.0	53.0	75.1	7.0	0.3	0.2
Avg	245.9	238.8	68.6	101.5	15.8	3.0	2.1
SD	157.3	157.7	11.1	10.9	8.9	1.8	1.1
Var	0.6	0.7	0.2	0.1	0.5	0.5	0.5

Furthermore, a Pearson correlation analysis was performed on the entire dataset comprising 544 samples to assess the linear relationships between the input variables (GGBS, FA, NaOH, Na_2SiO_3, and Age) and the target mechanical properties Ff and Ft. As illustrated in Figure 1, GGBS demonstrates a strong positive correlation with both Ff (0.77) and Ft (0.75), Indicating that higher levels of GGBS content are strongly associated with improvements in these strength characteristics. In contrast, FA, NaOH, and Na_2SiO_3 exhibit very weak correlations (near zero) with the output variables, indicating minimal direct linear influence on strength. Notably, the curing Age also shows moderate positive correlations with Ff (0.32) and Ft (0.30), reaffirming the critical role of hydration time in strength development. Most significantly, the interdependence between Ff and Ft is extremely high, which reflects their strong linear relationship as both are mechanical indicators of concrete performance.

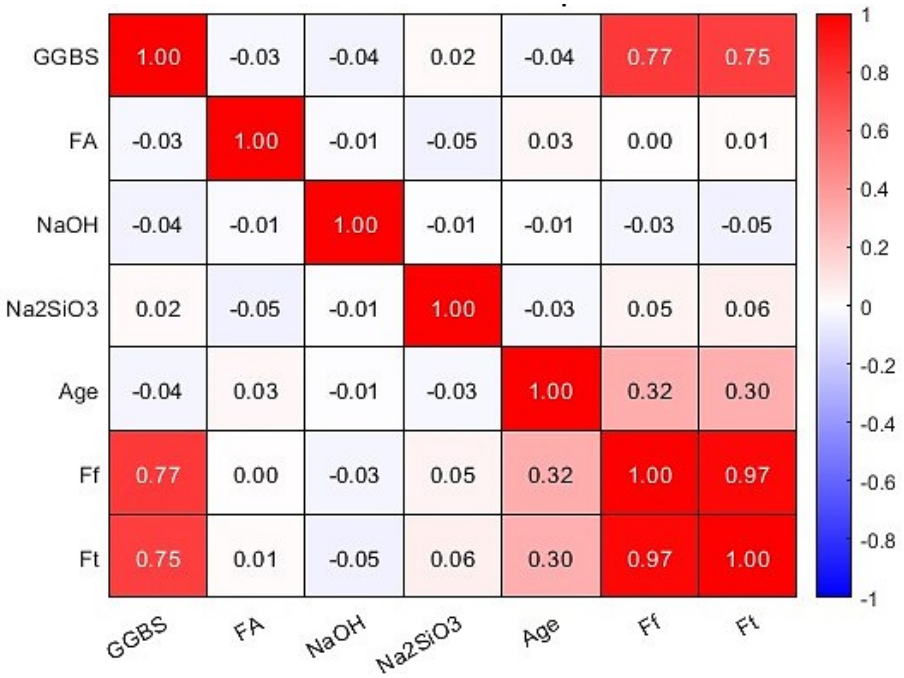


Figure 1. Correlation heatmap illustrates the Pearson correlation coefficients among input variables and output properties Ff, and Ft

2.1.3. Suitability of Dataset Size and Modeling Considerations

The dataset used in this study comprises 544 records, which is considered moderately large for machine learning applications in the concrete domain. Prior studies have successfully trained reliable models on smaller datasets, often fewer than 500 samples, demonstrating that this size provides adequate variability and learning depth for predictive modeling [30,31]. For algorithms, this dataset size offers a good balance between training volume and prediction

accuracy. However, further data enrichment may be required for more complex deep learning architectures.

The study incorporates cross-validation and bootstrapping techniques to enhance model generalization and stability. Future improvements could include the application of data augmentation or synthetic data generation methods, especially if domain-specific edge cases or underrepresented mix designs are identified [32].

Figure 2 schematically presents the methodology flow, highlighting the sequence from data acquisition through preprocessing, training, evaluation, and validation. The integration of statistical checks, data balancing, and robust model validation techniques ensures the proposed framework's reliability and replicability.

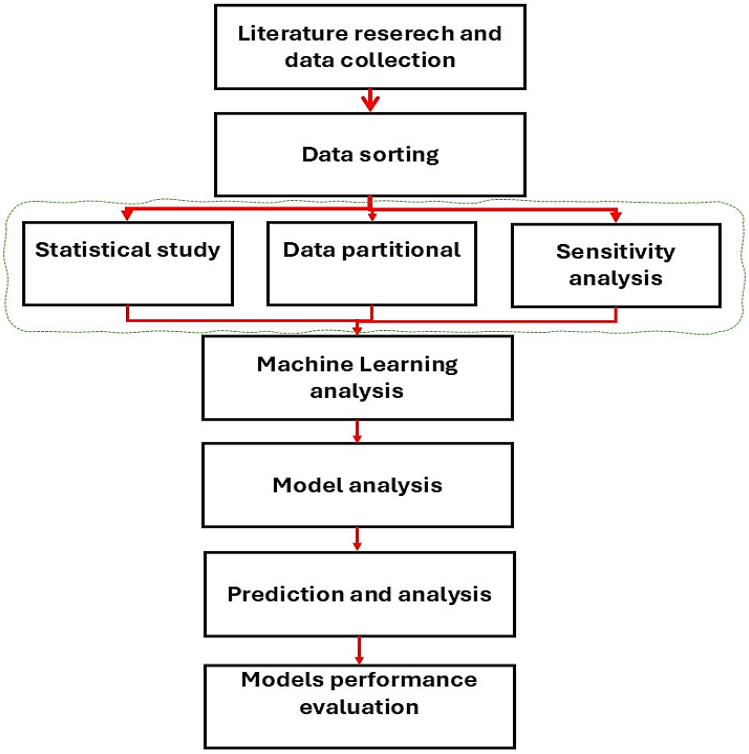


Figure 2. Workflow diagram

2.2. Machine Learning Framework

The selection of ML models plays a critical role in developing reliable predictive systems for estimating SCGC's mechanical behavior. This study employed eight supervised learning algorithms, including ensemble, rule-based, probabilistic, and kernel-based methods, to achieve both accuracy and interpretability.

2.2.1. Gradient Boosting (GB)

GB is an ensemble technique that constructs a series of weak models, commonly decision trees, where each model is trained to reduce the residual errors of the previous one. This step-by-step optimization process minimizes a specific loss function through gradient-based updates. The final predictive model is represented in Eq. (1).

$$y^{-}(x) = \sum_{m=1}^{M} \alpha_m h_m(x) \quad (1)$$

In the context of the GB model, the predicted output $y^{-}(x)$ is computed as the weighted sum of multiple weak learners, representing the final predicted value for a given input x. The

https://doi.org/10.31436/iiumej.v26i3.3730

term α_m enotes the weight assigned to the m[th] weak learner, reflecting its contribution to the overall prediction. The function $h_m(x)$ refers to the individual weak learner, typically a decision tree, trained during the m^{th} iteration. The model constructs a total of M such learners, each aimed at minimizing the residual error from the previous iteration, thus refining the prediction iteratively. This ensemble approach enables Gradient Boosting to capture complex, nonlinear relationships and enhance predictive accuracy.

Its strength lies in achieving high predictive performance for complex regression tasks while effectively handling linear and non-linear relationships. Figure 3 represents the Gradient Boosting framework applied to regression tasks. The process begins with the initial dataset (Database X), where the first decision tree $f_1(x, \theta_1)$ is trained. Subsequent trees (f_2, f_3,..., f_k) are iteratively constructed to predict the residuals from the previous trees. Each new tree corrects the errors of the ensemble up to that point. The final model output is the sum of predictions from all trees $\sum f_k(x, \theta_k)$, where an objective function guides node splitting to minimize prediction error.

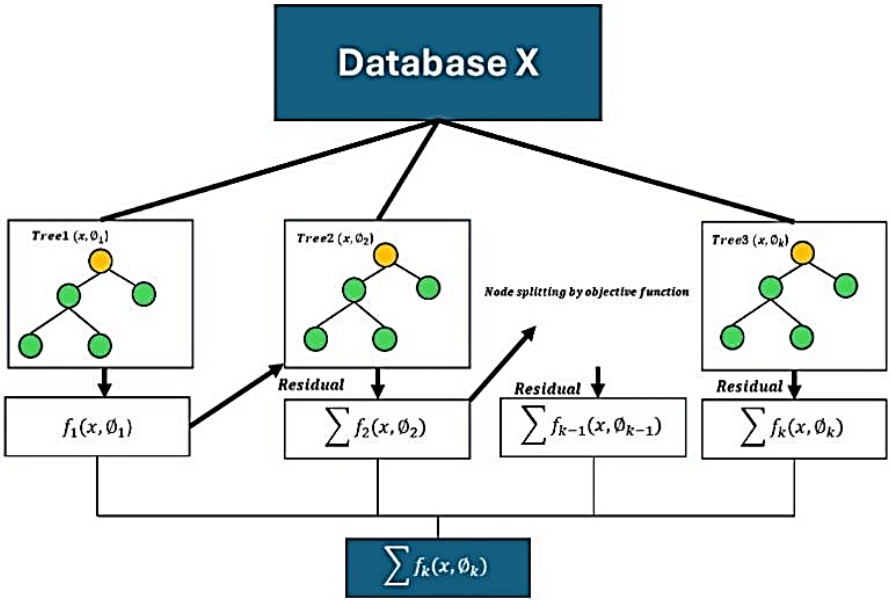


Figure 3. Schematic representation of the Gradient Boosting framework applied to regression tasks

2.2.2. CN2 Rule Induction

The CN2 algorithm is a rule-based learning method designed to produce understandable decision rules, often used for classification problems. It creates clear if-then statements by maximizing a quality measure that evaluates how effectively each rule represents the data [33]. A commonly used heuristic is in Eq. (2).

$$Score = \frac{Correct\ Predictions}{Total\ Predictions} \times \log\left(\frac{Correct\ Predictions}{Total\ Observations}\right) \quad (2)$$

CN2 supports categorical and numerical features and is appreciated for its transparency and comprehensibility [29]. Figure 4 shows the CN2 framework. The left side shows the construction of input behavior indicator sequences (S1 to Sn) from raw input entries, while the right side depicts the construction of output technical indicators (P1 to Pn). The final output is a set of predictions or classifications (C1 to CN), reflecting the model's learned patterns from both behavioral and technical indicators.

https://doi.org/10.31436/iiumej.v26i3.3730

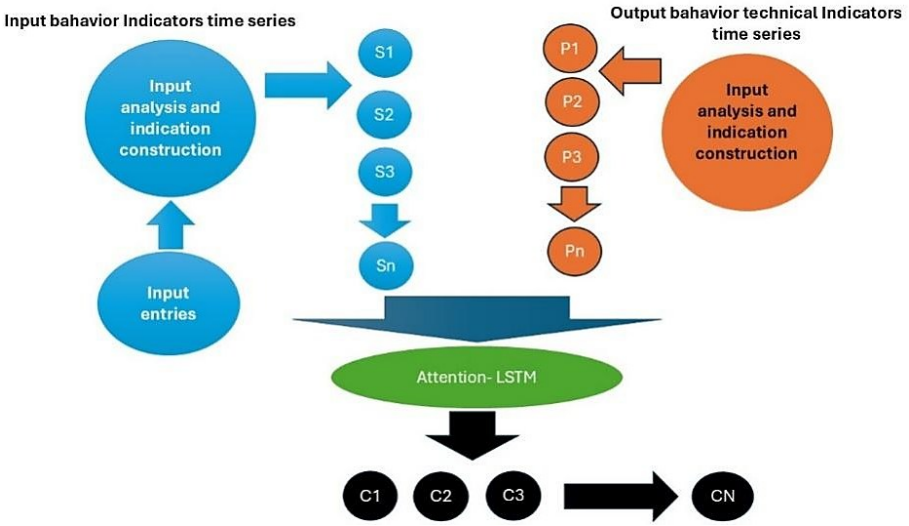


Figure 4. Diagram illustrating the Attention-LSTM framework for time-series modeling of input and output behavior indicators

2.2.3. Naive Bayes Classifier

Naive Bayes is a probabilistic model rooted in Bayes' Theorem, assuming conditional independence among features given the class label [34]. Despite its simplicity, it is highly effective for high-dimensional problems. The model estimates the posterior class probability as Eq. (3).

$$P(C|X) = \frac{P(X|C)\cdot P(C)}{P(X)} \tag{3}$$

where $P(C \mid X)$ is the probability of class C given feature vector X, $P(X \mid C)$ is the likelihood of X under class C, $P(C)$ is the prior class probability, $P(X)$ is the marginal probability of X. Figure 5 illustrates the Naïve Bayes classification process. The left side shows a dataset of mixed features (represented by different shapes and colors). The core of the process applies Bayes' Theorem, where the probability of class C given input X is computed based on prior and likelihood probabilities. The output on the right shows the data being classified into different categories (triangles, rectangles, circles), demonstrating the probabilistic sorting capability of the Naïve Bayes model.

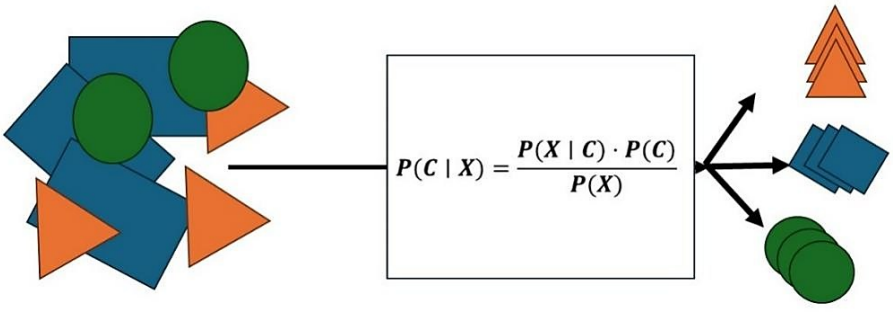


Figure 5. Illustration of the Naïve Bayes classification process

2.2.4. Support Vector Machines (SVM)

Support Vector Machines seek to construct a hyperplane that maximizes the margin between classes in a high-dimensional space [35]. For a linear SVM, the optimization problem is defined as Eq. (4).

https://doi.org/10.31436/iiumej.v26i3.3730

$$min\ \frac{1}{2}\ ||\ w\ ||^2\ \text{ subject to } y_i(w^t x_i + b) \geq 1, \forall_i \quad (4)$$

where w is the weight vector, x_i is the input feature vector, y_i is the corresponding class label, b is the bias term.

SVMs are well-suited for structured data and offer robust performance with different kernel functions. Figure 6 shows the architecture of an SVR model. On the left, the SVR network structure is depicted with an input layer receiving feature vectors $(x_1, x_2, \ldots, x_n)$, which are transformed through a kernel function $k(x_i,x)$ in the hidden layer. These kernel evaluations are aggregated in the output layer along with a bias term b, resulting in the predicted output y. The right side shows the concept of univariate linear SVR, where the goal is to fit a linear function $f(x) = wx + b$ such that the deviations from actual values lie within a defined margin ϵ. Points outside this margin incur an error penalty ε, forming the basis for the SVR loss function.

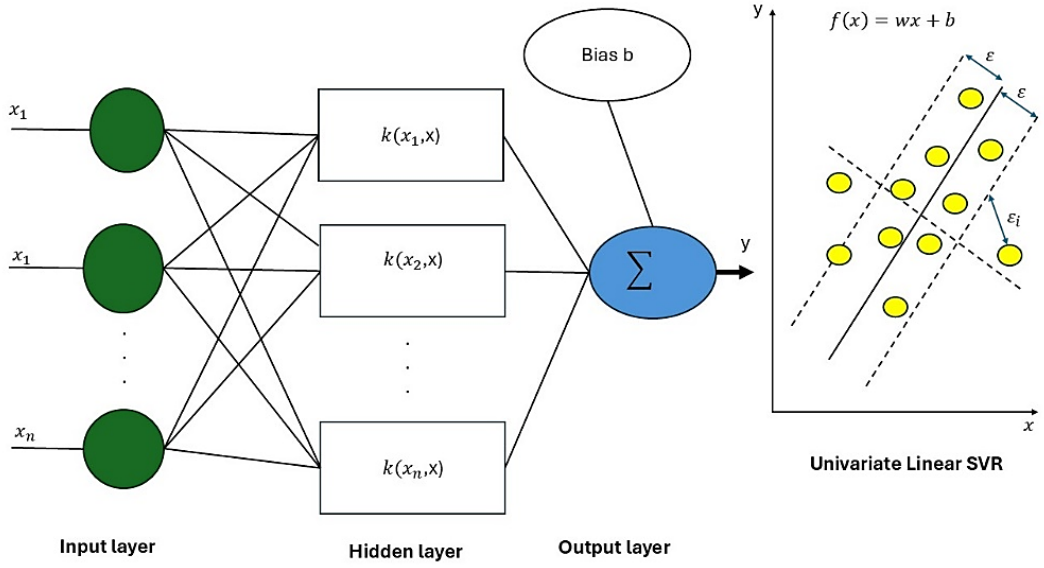


Figure 6. Architecture of a Support Vector Regression (SVR) model

2.2.5. Stochastic Gradient Descent (SGD)

SGD is an optimization algorithm used to minimize the loss function by iteratively updating model parameters based on gradients from small batches or individual training samples [36]. It is formulated as Eq. (5).

$$\Theta = \theta - \eta \nabla L\left(\theta; x_{ily_i}\right) \quad (5)$$

where θ is model parameters, η is learning rate, ∇L is gradient of the loss function with respect to θ.

SGD is especially useful in large-scale or streaming data scenarios due to its computational efficiency [29, 30]. Figure 7 visualizes the gradient descent optimization process used in machine learning model training. The graph shows the relationship between model weight and loss reduction. Starting from an initial weight, the model updates its parameters toward the negative gradient, iteratively minimizing the loss function. The global minimum is achieved when the gradient becomes zero, indicating optimal model parameters with the lowest possible loss.

https://doi.org/10.31436/iiumej.v26i3.3730

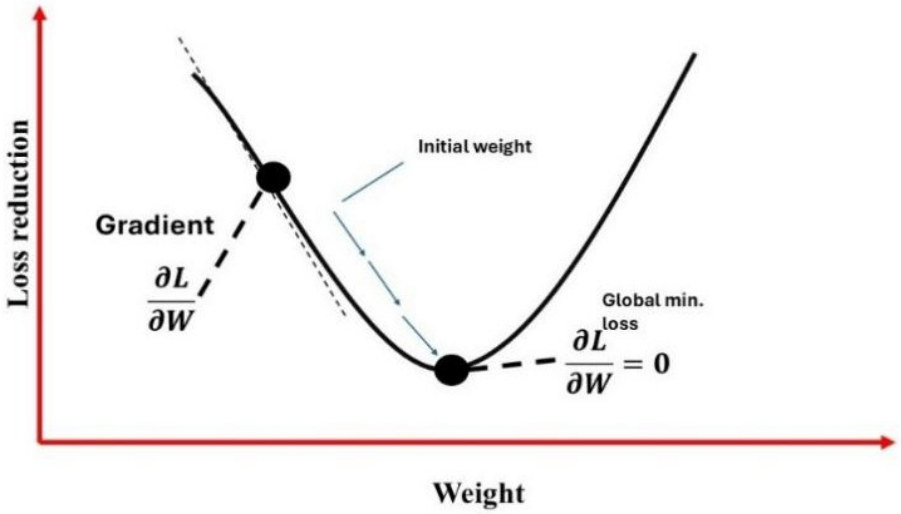


Figure 7. Gradient descent optimization process used in machine learning model training

2.2.6. K-Nearest Neighbors (KNN)

KNN is a non-parametric, instance-based learning method. It predicts the output based on the majority vote (classification) or average (regression) of the k-closest training samples [37] represented by Eq. (6).

$$F(x) = \frac{1}{k}\sum_{i \in kNN} y_i \tag{6}$$

where y_i is the labels of the nearest neighbors. KNN is intuitive and requires no training phase, but performance can be sensitive to feature scaling and distance metrics. Figure 8. Illustration of the K-Nearest Neighbors (K-NN) classification process. The left panel ("Before") shows a new data point introduced into a feature space defined by variables, surrounded by known samples from Category A and Category B. Using the K-NN algorithm, the right panel ("After") depicts how the new data point is classified based on its proximity to the majority.

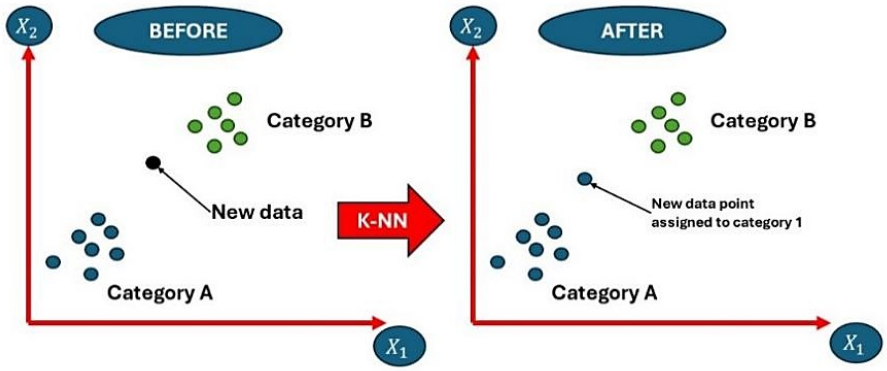


Figure 8. K-Nearest Neighbors (K-NN) classification process

2.2.7. Decision Trees (DT)

Decision Trees are recursive models that partition the data based on feature splits that maximize information gain or minimize impurity [38]. The Gini Index is a common criterion used for classification tasks represented in Eq. (7).

$$Gini = 1 - \sum_{i=1}^{C} p_i^2 \tag{7}$$

https://doi.org/10.31436/iiumej.v26i3.3730

where p_i is the proportion of class i in a node, and C is the number of classes. Trees are simple to interpret and can model non-linear relationships effectively. Figure 9 is a structure of a decision tree model. The diagram illustrates the hierarchical flow of decisions starting from the root node (top red circle), followed by multiple split nodes (intermediate red circles) representing decision rules based on input features. The process continues until reaching the leaf nodes (yellow circles), which represent final predictions or class labels. This structure enables the model to partition the data space for classification or regression tasks recursively.

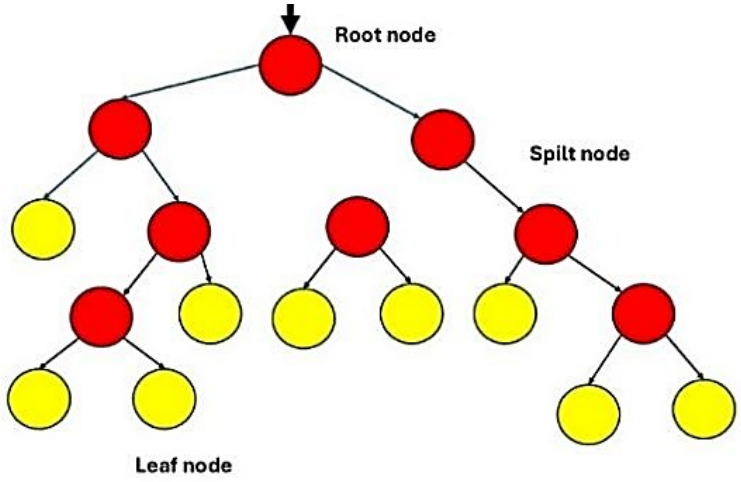


Figure 9. Structure of a decision tree model

2.2.8. Random Forest (RF)

Random Forest is an ensemble of decision trees built on bootstrapped samples of the dataset. Predictions are aggregated through majority voting or averaging [39]. The ensemble output is given by Eq. (8):

$$y^{-} = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \tag{8}$$

where T is the number of trees, $h_t(x)$ output from the t^{th} tree.

Random Forest improves generalization by reducing overfitting, a common issue in single decision tree models. Figure 10. Illustration of a decision tree using logical conditions for binary classification. The tree begins with a root condition on $Variable_a \geq x$. If true, the data point is classified as TRUE (green node); if false, the tree proceeds to evaluate $Variable_b \leq y$. Further branching leads to additional conditions until a final classification of TRUE or FALSE (green or red leaf node) is reached. This logical flow demonstrates how decision rules are applied in a tree-based model to reach categorical outcomes.

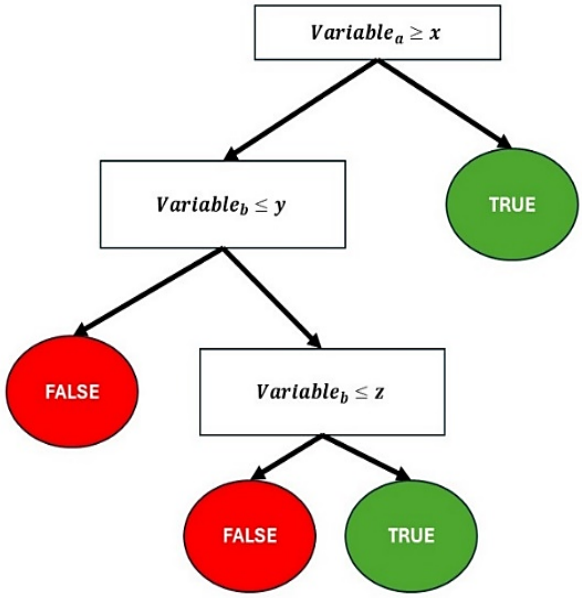


Figure 10. Decision tree using logical conditions for binary classification

The choice of algorithm in this study was guided by theoretical characteristics and empirical performance in modeling SCGC properties. Hyperparameter tuning was conducted using grid search and cross-validation to achieve optimal configurations. These algorithms offer a spectrum of benefits, ranging from interpretability (e.g., CN2, Decision Trees) to high accuracy and scalability (Gradient Boosting, Random Forest, SVM), thereby providing a robust framework for predictive analysis of SCGC mechanical behavior.

2.3. Response Surface Methodology (RSM), Hoffman & Gardener, and Analysis of Variance (ANOVA)

RSM statistical use for modeling and optimizing processes in which multiple input variables influence a quantitative outcome or response. It provides a systematic methodology for understanding the interactions among process variables and for determining the optimal operating conditions of complex systems [40]. RSM constructs an empirical model, commonly a second-degree polynomial, that approximates the actual relationship between the independent variables and the output response, as expressed in Eq. (9).

$$Y = \beta_0 + \sum \beta_i Xi + \sum \beta_{ii} X_i^2 + \sum \beta_{ij} X_i X_j + \varepsilon \quad (9)$$

where Y is the response variable, X_i are the input factors, β terms are the regression coefficients, and ε represents the experimental error [4].

In this study, RSM was incorporated alongside ML models (KNN and SVM to leverage the strengths of both analytical and data-driven approaches. While ML algorithms are highly effective for capturing complex, non-linear patterns in experimental data, they often lack transparency regarding how input variables influence the response. RSM addresses this gap by offering an interpretable, regression-based framework that models variable interactions and supports optimization through structured mathematical formulations. Therefore, the integration of RSM is not aimed at outperforming ML in predictive capacity, but rather at enhancing the interpretability, parametric sensitivity analysis, and optimization of mix design parameters. As illustrated in Figure 11, the accompanying decision-making diagram, both approaches operate on the same dataset but fulfill complementary roles. RSM provides equation-based insights and

design guidance, while ML supports robust pattern recognition and generalization across a wide input space.

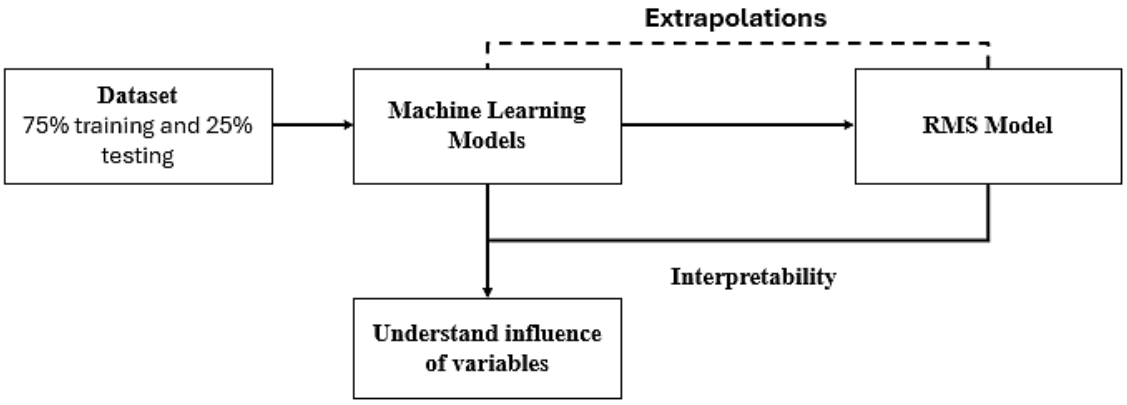


Figure 11. Methodological framework showing the complementary roles of RSM and machine learning

ANOVA is a statistical method used to break down the overall variability in a response variable into separate components, such as those caused by experimental conditions and inherent randomness. In the context of RSM, ANOVA plays a key role in evaluating the significance and adequacy of the regression model developed [41]. It tests the null hypothesis that the means of various factor levels are equal and provides an F-ratio to determine if the regression model explains a significant portion of the variance in the data. The formula is as in Eq. (10)

$$F = \frac{Mean\ Square\ Regression\ (MSR)}{Mean\ Square\ Error\ (MSE)} \tag{10}$$

A low p-value (typically < 0.05) associated with the F-statistic indicates that the model terms significantly contribute to the response prediction. ANOVA also provides insight into the contribution of linear, interaction, and quadratic terms, allowing researchers to refine the model structure by retaining only statistically significant terms [41].

RSM and ANOVA form a robust and complementary framework for prediction and optimization in experimental research. RSM enables the construction of mathematical models that capture how input parameters influence the output response, whereas ANOVA provides a means to statistically verify the model's validity, effectiveness, and predictive strength This integrated approach enhances decision-making by enabling researchers to identify the main and interaction effects of key process variables, assess model goodness-of-fit and lack-of-fit, and determine optimal factor combinations for achieving target performance outcomes. Moreover, by systematically exploring the experimental space, the RSM–ANOVA combination reduces the required trials, maximizing information gain with minimal resource use. In the context of materials science, particularly in the formulation of SCGC, this dual methodology proves especially valuable for fine-tuning mix designs and curing conditions to achieve superior mechanical properties cost- and time-efficiently.

The same 75% training and 25% testing data split was consistently applied to all modeling approaches, including the machine learning algorithms and the RSM models, to ensure a fair and unbiased comparison across modeling techniques. This uniform data partitioning strengthens the validity of comparative performance evaluations and mitigates potential biases introduced by differing datasets.

In addition to RSM and ANOVA, a sensitivity analysis was conducted using the Hoffman & Gardener method to evaluate how each input variable affects the predicted mechanical performance outcomes. Although not a standard sensitivity approach in concrete modeling, this method has been successfully applied in recent machine learning studies involving concrete composites [42]. It involves systematically perturbing one input at a time while keeping the others constant and evaluating the resulting change in the output prediction from trained models. [43].

Although the method involves considerable computational effort, it offers significant benefits in complex modeling tasks where identifying the influence of input variables is essential for optimization and informed decision-making. In this research, an initial sensitivity investigation was performed using the developed dataset to assess how each independent variable contributes to the outcome Y. A one-factor-at-a-time approach was applied, in which a single input is varied. In contrast, all others remain fixed, enabling a focused evaluation of individual variable influence. The Sensitivity Index (SI) for each input was computed using the Hoffman & Gardner technique [44], defined by the following Eq. (11).

$$SI(X_N) = \frac{Y(X_{max}) - Y(X_{min})}{Y(X_{max})} \tag{11}$$

A sensitivity value equal to 1.0 signifies maximum responsiveness of the output to that input variable. In contrast, a value below 0.01 reflects negligible influence, suggesting that the model's predictions remain largely unaffected by variations in that parameter.

2.4. Performance Accuracy Evaluation of Machine Learning Models

A set of statistical performance metrics was employed to assess the predictive capabilities of the developed ML models for estimating the mechanical properties of SCGC.

2.4.1. Error-Based Metrics

Sum of Squared Errors (SSE) quantifies the total squared deviation between actual and predicted values, serving as a cumulative indicator of prediction error. While useful for comparative analysis, SSE is scale-dependent and less intuitive when comparing across datasets [45]. MAE (Mean Absolute Error) measures the average magnitude of prediction errors, offering a direct and easily interpretable estimate of how far predictions deviate from actual values [46] represented by Eq. (12).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - y_i^-| \tag{12}$$

The Mean Squared Error (MSE) is a widely used metric for evaluating the accuracy of predictive models. It is calculated by squaring the prediction errors (the differences between observed and predicted values) and then averaging these squared errors. This squaring process amplifies the influence of larger deviations, making MSE particularly sensitive to outliers [47]. The formula represented by Eq. (13).

Root Mean Squared Error (RMSE) is a widely used metric for evaluating the performance of models, particularly in fields like hydrology, geoscience, and machine learning. calculates the square root of the mean squared deviations between actual and predicted values, offering a concise indicator of overall prediction accuracy [48]. The formula represented by Eq. (14).

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - y_i^-)^2 \tag{13}$$

https://doi.org/10.31436/iiumej.v26i3.3730

$$RMSE = \sqrt{MSE} \tag{14}$$

These metrics (MAE, MSE, $RMSE$) provide complementary perspectives on error behavior and are essential for understanding average and extreme prediction performance.

2.4.2. Normalized Performance Measures

Error (%) expresses prediction error as a percentage relative to the predicted value, enabling model comparisons across variables or datasets with different scales. The formula is represented by Eq. (15). Accuracy (%) is calculated as the complement of error, representing the proportion of accurate predictions in percentage terms. The formula is represented by Eq. (16).

$$\text{Error}(\%) = \frac{\text{RMSE}}{\bar{y_i}} \times 100 \tag{15}$$

$$\text{Accuracy}(\%) = 100 - \text{Error} \tag{16}$$

Although frequently associated with classification tasks, Error (%) and Accuracy (%) can also be adapted for use in regression settings to provide scale-independent performance indicators.

2.4.3. Coefficient of Determination (R^2)

R^2 indicates the proportion of variance in the observed data that the model explains. It ranges from 0 to 1, with values closer to 1 indicating a better fit. Unlike absolute error metrics, R^2 evaluates model goodness-of-fit in terms of variance reduction, and the formula represented by Eq. (17).

$$R^2 = 1 - \frac{\sum(y_i - \hat{y_i})^2}{\sum(y_i - \bar{y})^2} \tag{17}$$

Here, y_i are the observed values, $\hat{y_i}$ are the predicted values, and $\bar{y}$ is the mean of observed values. A high R^2 implies that the model captures the underlying patterns of the data effectively [49].

2.4.4. Multi-Metric Evaluation Strategy

Given the limitations and strengths of each metric, a multi-metric evaluation framework was adopted. For instance, a model may achieve a high R^2 while yielding large MAE or RMSE if influenced by outliers. Similarly, low error percentages do not necessarily guarantee high explanatory power if model predictions fail to capture data variance. Therefore, combining these metrics enables a balanced and comprehensive assessment of performance. This performance evaluation framework ensures that the selected ML models for SCGC prediction are statistically robust and practically relevant across varying strength parameters.

3. RESULTS AND DISCUSSION

3.1. Model Performance

3.1.1. Gradient Boosting (GB) model

GB models were implemented using the Scikit-learn library. The models employed a constant rate of 0.1 and a minimum sample split threshold of 2. A series of nine model setups

https://doi.org/10.31436/iiumej.v26i3.3730

was tested by progressively varying the number of decision trees (from 1 to 3) and their corresponding depth levels (from 1 to 3). The influence of these parameters on prediction error is illustrated in Figure 12, where a consistent decrease in error percentage is observed with greater model complexity. Specifically, increasing tree depth and count significantly enhanced model accuracy, particularly in reducing prediction error for Ff and Ft.

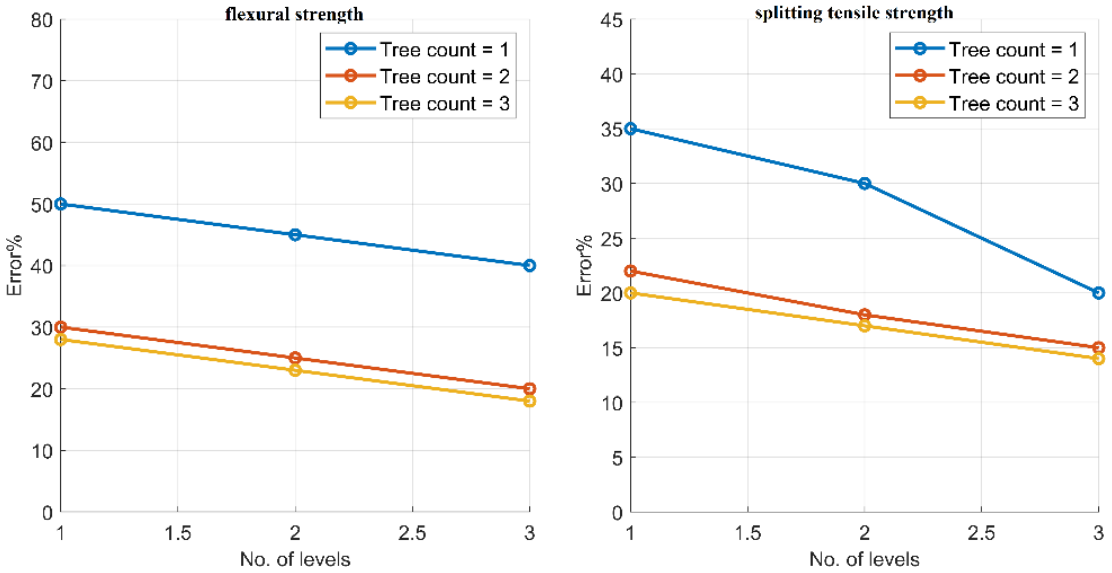


Figure 12. Gradient Boosting (GB) model performance for predicting Ff and Ft

The optimal GB model configuration was identified as two trees with a maximum depth of three levels. Under this configuration, the GB models achieved average prediction accuracies of 90% for Ff and 88% for Ft. Figure 13 and Figure 14 illustrate the correlation between the predicted strength values and the corresponding experimental results. In the training phase, the model attained an R^2 of 0.97 for Ff and 0.95 for Ft, with a near-ideal regression slope (y = 0.99x), indicating high predictive alignment. In the validation phase, the R^2 remained robust at 0.96, affirming the model's generalization capability and reliability in estimating the tensile-related properties of SCGC.

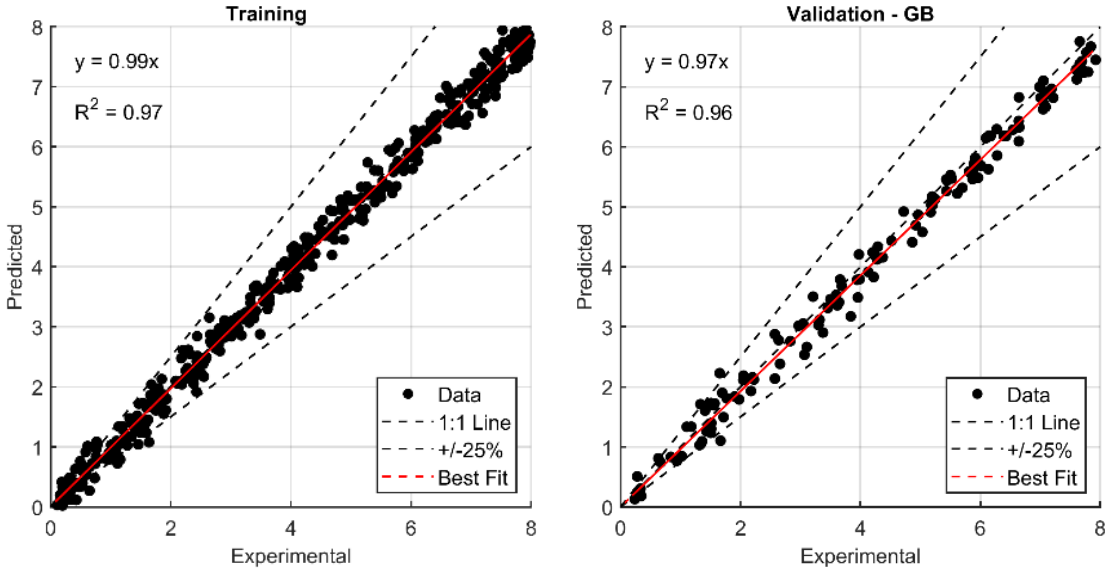


Figure 13. Regression analysis comparing predicted and experimental values for Ff using the Gradient Boosting model

https://doi.org/10.31436/iiumej.v26i3.3730

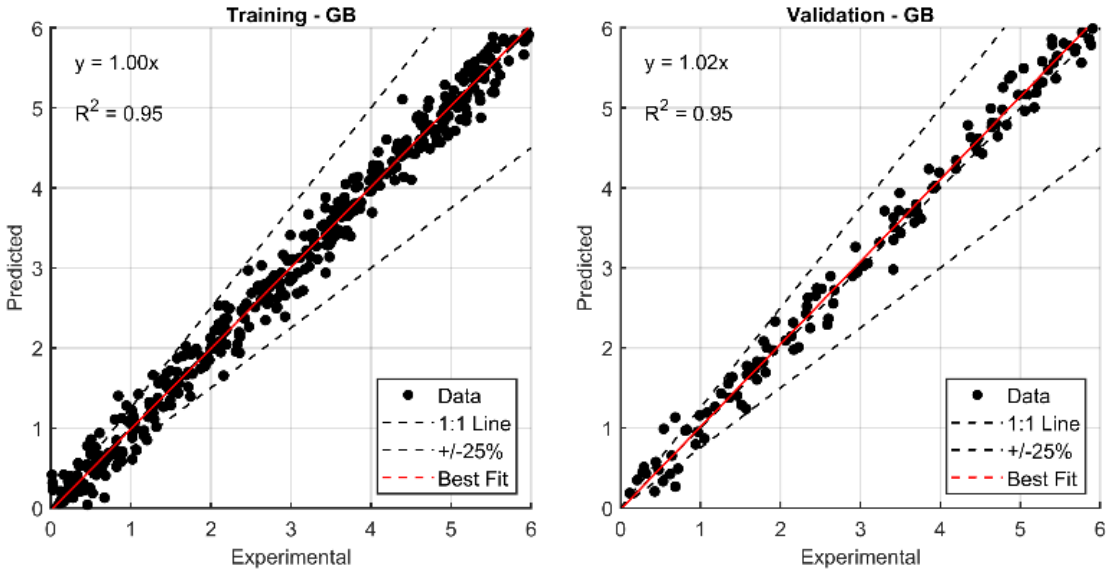


Figure 14. Regression analysis comparing predicted and experimental values for Ft using the Gradient Boosting model

3.1.2. CN2 model

Nine CN2-based rule induction models were constructed to model the mechanical performance of SCGC, Ff, and Ft. The assessment was based on Laplace accuracy, using a beam width of 1.0 and setting the minimum coverage for rule generation to 1.0. Model refinement was done by gradually extending the maximum rule length from 1 to 5. As depicted in Figure 15, extending the rule length consistently improved prediction accuracy, with the configuration at a length of 5.0 yielding the lowest error rates for both target properties.

Each developed CN2 model generated a comprehensive set of interpretable "IF–THEN" rules 42 for Ff, and 31 for Ft. These rules captured complex relationships between the input mix design parameters and the corresponding strength outcomes.

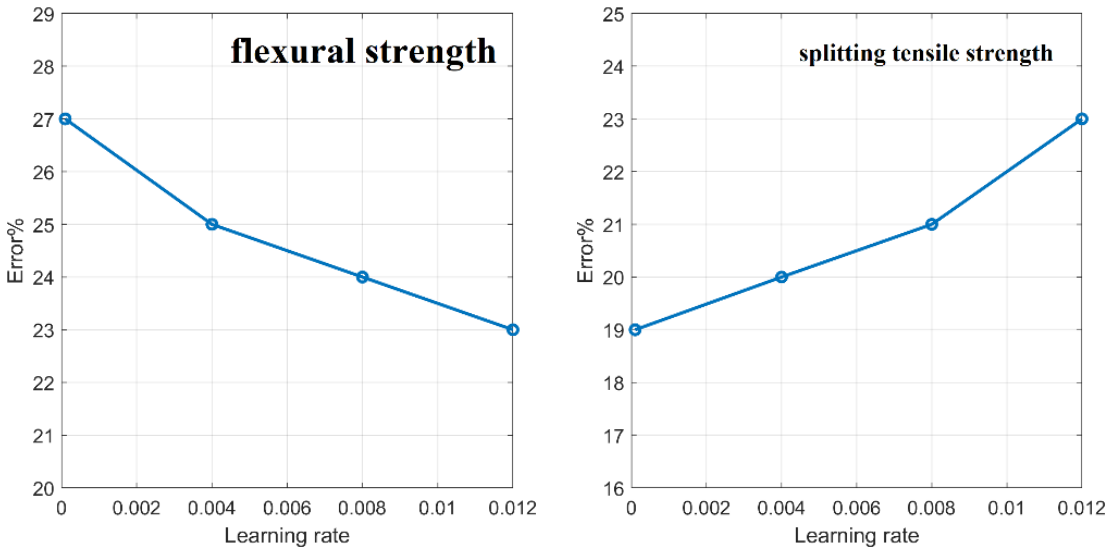


Figure 15. Error (%) reduction in predicting Ff and Ft using the CN2 model as the rule length increases from 1 to 5

https://doi.org/10.31436/iiumej.v26i3.3730

The regression plots in Figs. 16 and 17 demonstrate the close agreement between the predicted and experimental strength values across all three strength types. For Ff and Ft, the training and validation phases achieved high coefficients of determination ($R^2 \approx 0.98$ and 0.94), with nearly ideal regression slopes. These results validate the effectiveness of the CN2 model in capturing the governing patterns of SCGC mechanical behavior, while offering transparent decision rules for practical implementation in sustainable concrete mix design.

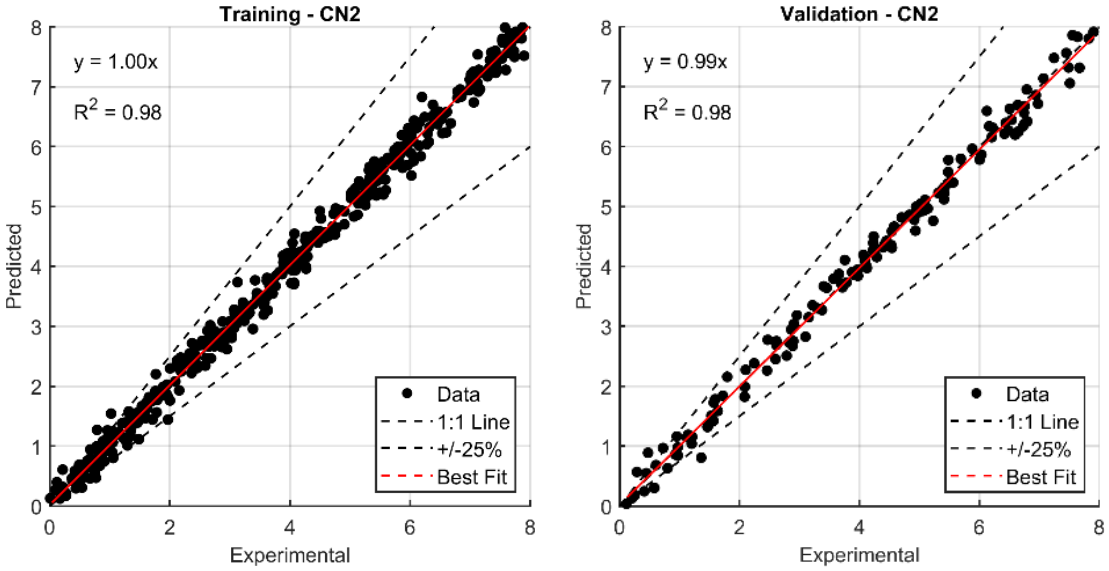


Figure 16. Regression analysis of CN2 model predictions versus experimental results for flexural strength

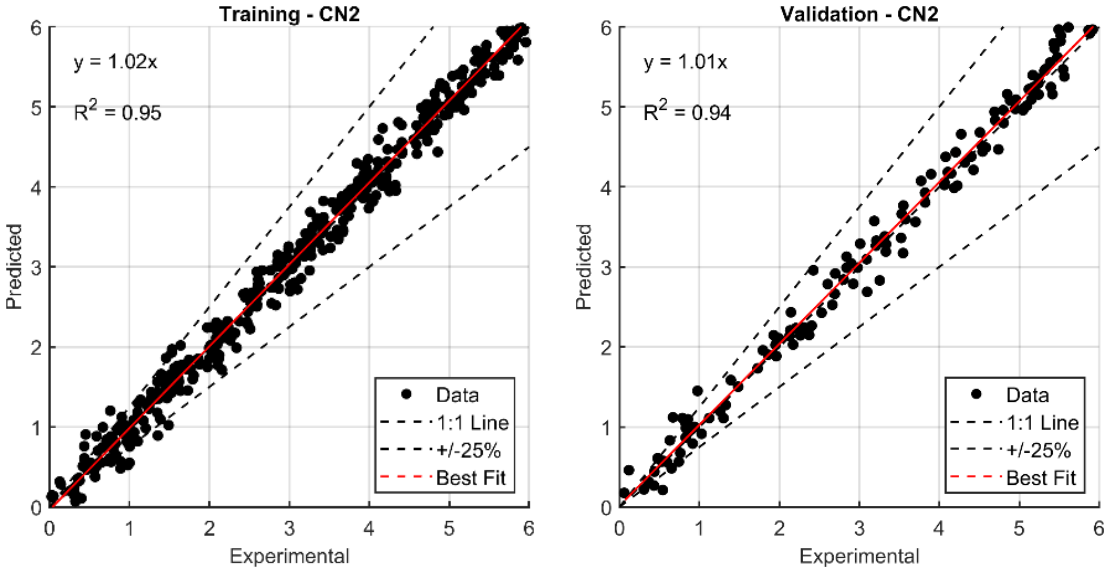


Figure 17. Regression analysis of CN2 model predictions versus experimental results for splitting tensile strength

3.1.3. Support Vector Machine (SVM) model

SVM models were employed to estimate the mechanical properties of SCGC, using a polynomial kernel function. Model optimization was performed by assigning a regularization constant (C) of 100, setting the loss margin for regression to 0.10, and applying a convergence threshold of 1.0. Polynomial degrees from 1 (linear) to 3 (cubic) were methodically explored

https://doi.org/10.31436/iiumej.v26i3.3730

across multiple configurations to assess how complexity influences predictive performance. As shown in Figure 18, increasing the kernel degree led to a noticeable reduction in prediction error for flexural and tensile strength, with the quadratic kernel (degree 2) providing the best balance between complexity and accuracy.

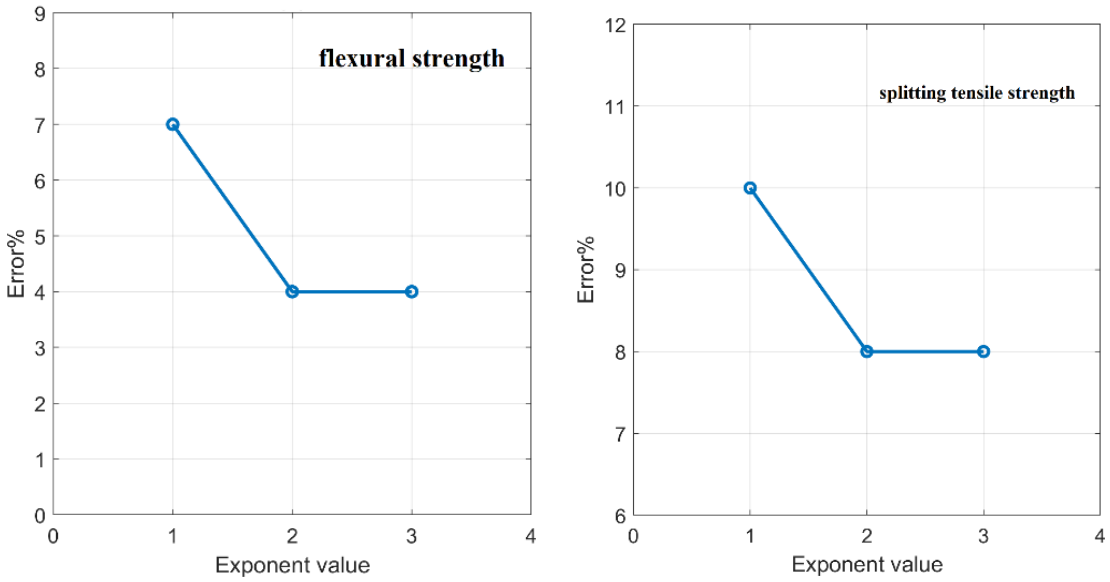


Figure 18. Prediction error (%) for Ff and Ft using SVM models with polynomial kernels of increasing degree

The SVM models demonstrated excellent predictive capacity across all target properties, with average accuracies of 99% and 99% for Ff and Ft, respectively. Regression plots in Figs. 19 and 20 confirm the strong correlation between experimental and predicted values. The models achieved R^2 values of up to 0.99 for each strength property, with regression lines closely aligned with the ideal 1:1 line. These results affirm the robustness and high precision of the SVM models in estimating SCGC strength performance, particularly when using a second-degree polynomial kernel.

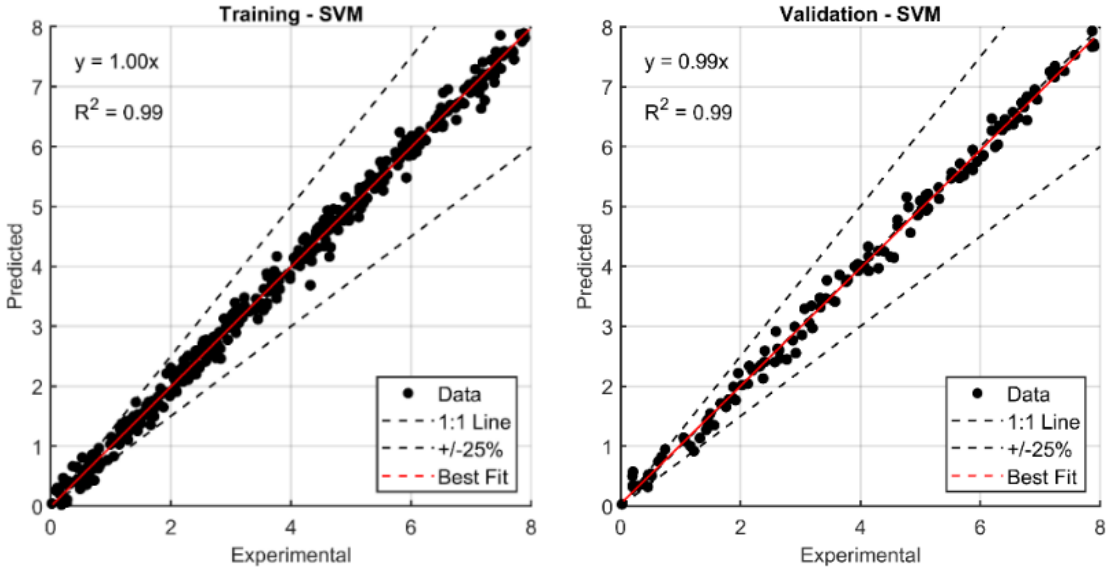


Figure 19. Regression plots comparing predicted and experimental results for Ff using the optimized SVM model (quadratic kernel)

https://doi.org/10.31436/iiumej.v26i3.3730

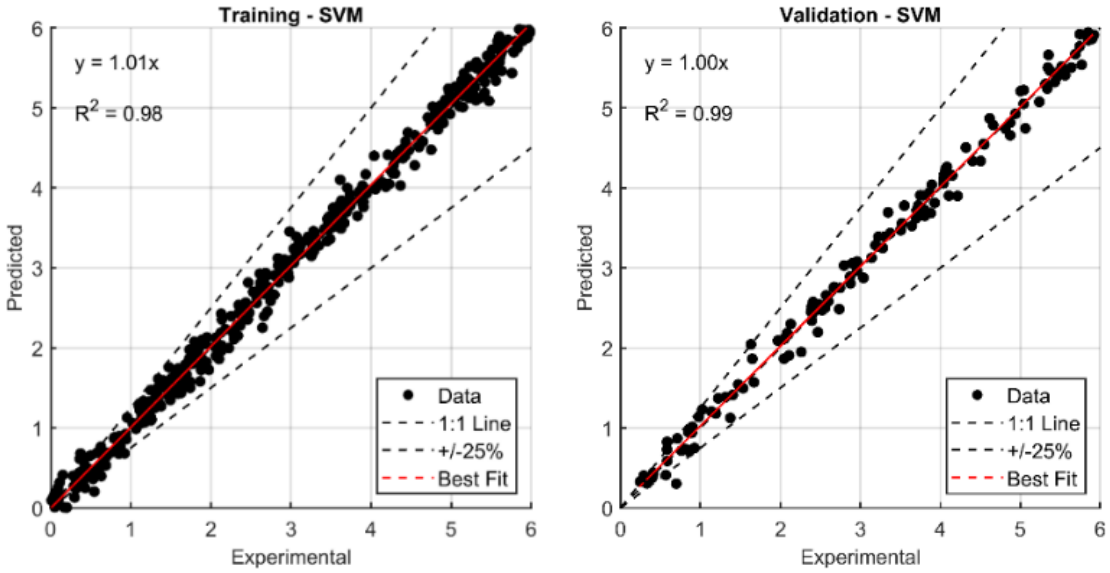


Figure 20. Regression plots comparing predicted and experimental results for Ft using the optimized SVM model (quadratic kernel)

3.1.4. Stochastic Gradient Descent (SGD model results

SGD models were developed using the modified Huber loss and elastic net regularization, with a blending ratio set at 0.01 and a regularization coefficient of 0.001. The models were evaluated across a learning rate range that began at 0.01 and gradually decreased to 0.001. As shown in Figure 21, learning rate adjustments had a significant impact on prediction error, with the lowest error observed at the smallest learning rate tested.

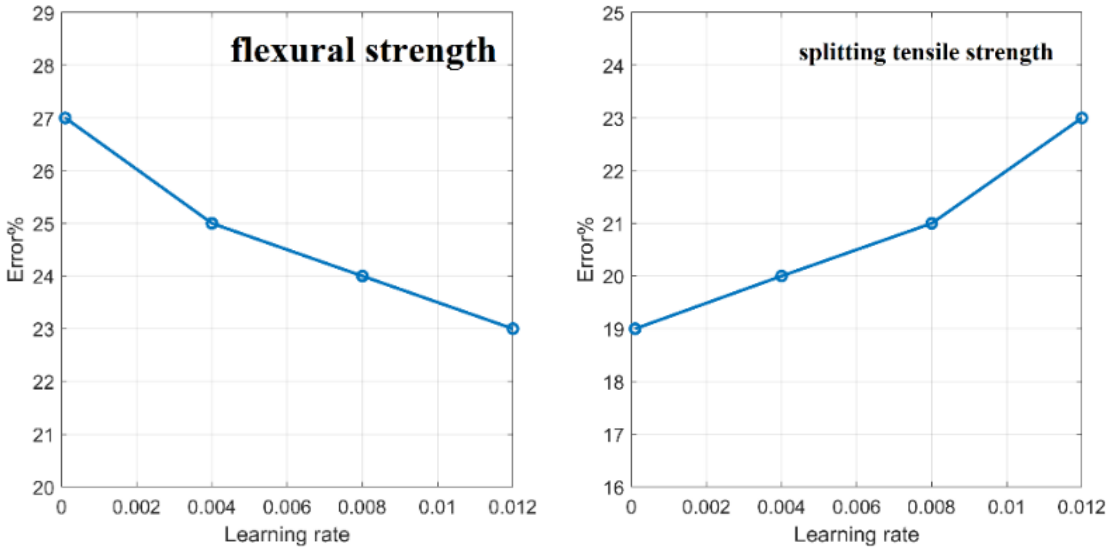


Figure 21. Effect of learning rate on prediction error (%) for Ff and Ft using the SGD model

Nine SGD model configurations were analyzed to estimate flexural and splitting tensile strength. Figure 21 and 22 present regression plots of predicted versus experimental values for both training and validation phases. The R^2 values ranged between 0.89 and 0.93 for flexural and tensile strengths. The results confirm the model's moderate yet consistent performance across the three strength categories, with good alignment along the 1:1 prediction line.

https://doi.org/10.31436/iiumej.v26i3.3730

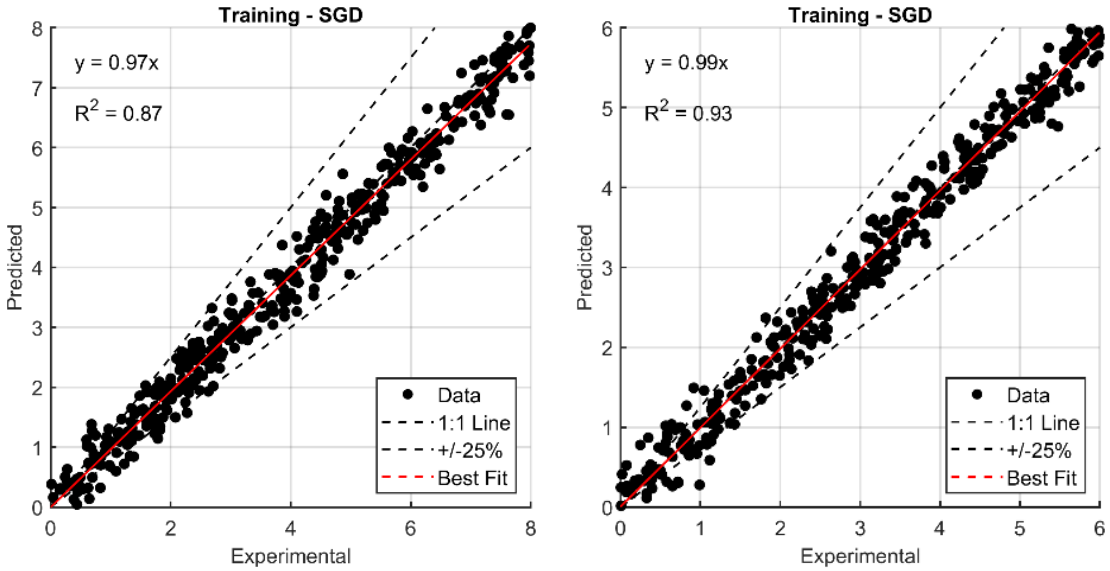


Figure 22. Predicted vs. experimental plots for Ff using the optimized SGD model

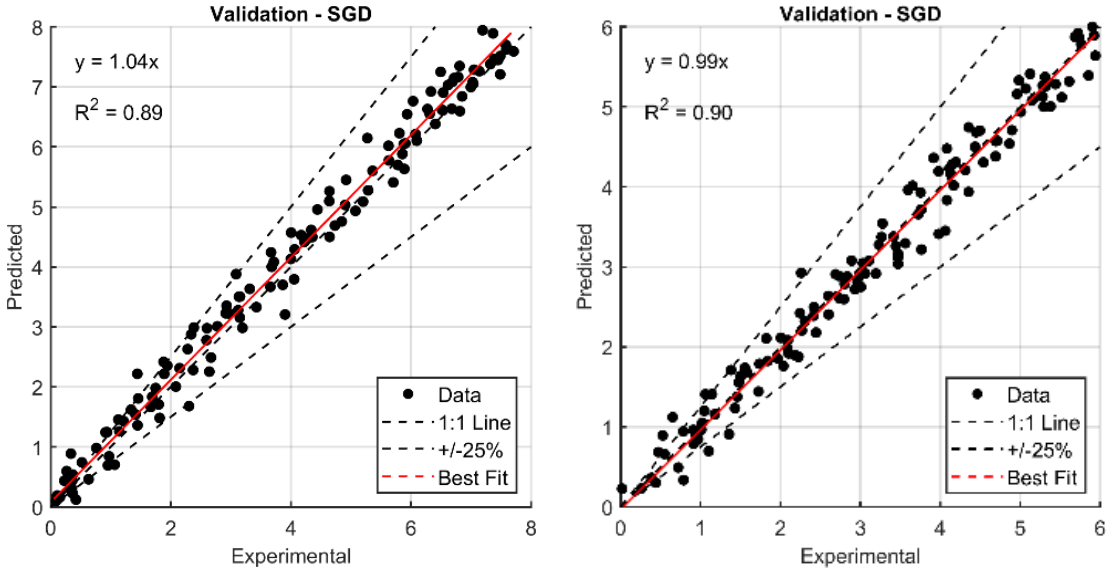


Figure 23. Predicted vs. experimental plots for Ft using the optimized SGD model

3.1.5. K-Nearest Neighbors (K-NN) model results

The K-Nearest Neighbors (K-NN) models were developed to predict the flexural and splitting tensile strength of SCGC. The optimal model configuration was determined using a single neighbor ($k = 1$) and the Euclidean distance metric, where each prediction was weighted according to the inverse of the distance from the query point. This setup allows the model to prioritize the most similar past observations, making it highly responsive to local patterns within the dataset. K-NN exhibited the highest predictive performance among all the machine learning algorithms evaluated in this study. The average prediction accuracies reached 99% for Ff, and 94% for Ft, indicating the model's exceptional capability in capturing the nonlinear interactions among input variables, as in Figs. 24 and 25. Its non-parametric and instance-based learning nature makes it particularly suitable for datasets with high variability and strong local dependencies, such as those found in SCGC mix designs.

https://doi.org/10.31436/iiumej.v26i3.3730

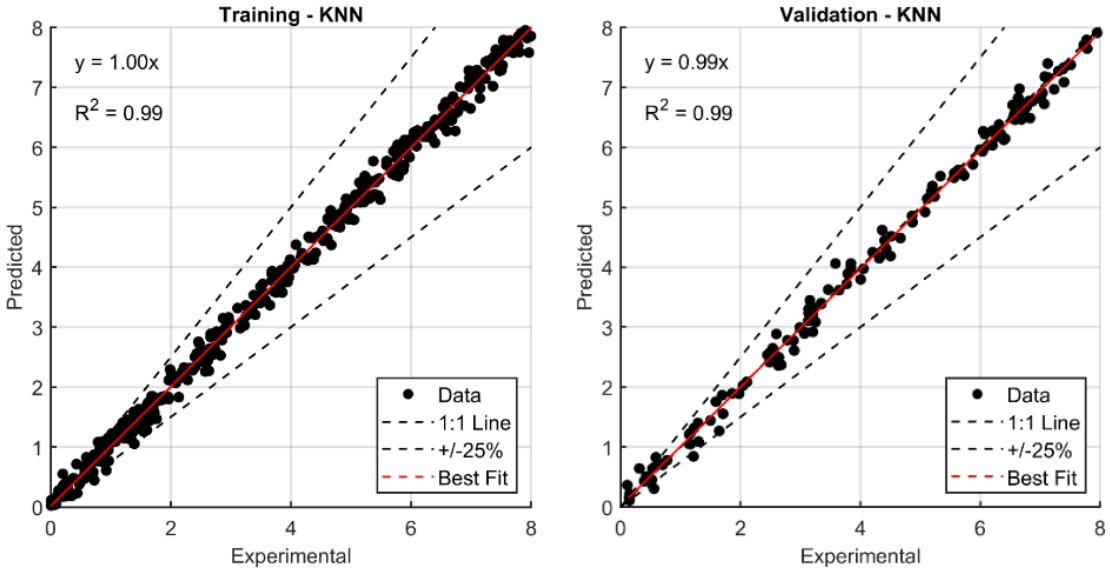


Figure 24. Predicted vs. experimental plots for Ft using the optimized K-NN model

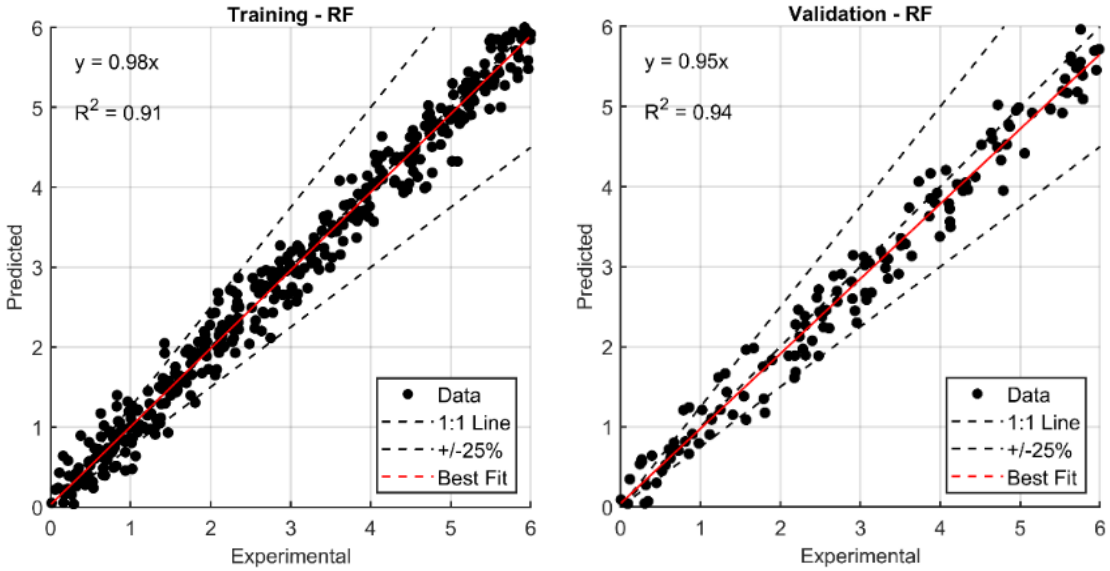


Figure 25. Predicted vs. experimental plots for Ft using the optimized K-NN model

3.1.6. Decision Tree (DT) model

DT models were built using a tiered node-splitting method. Each model was designed with at least 2.0 instances per leaf and a minimum split requirement of 5.0, balancing clarity of the decision rules with control over overfitting. Tree depth gradually increased, starting from a simple structure with one level and expanding up to five levels. This progressive layering allowed the algorithm to balance model complexity and predictive accuracy. The impact of tree depth on model performance is shown in Figure 26, where a consistent reduction in prediction error (%) is observed for both Ff and Ft as the number of layers increases. The most notable performance gains occur between depths 1 and 3, with diminishing returns beyond this point, indicating an optimal complexity level near five layers.

https://doi.org/10.31436/iiumej.v26i3.3730

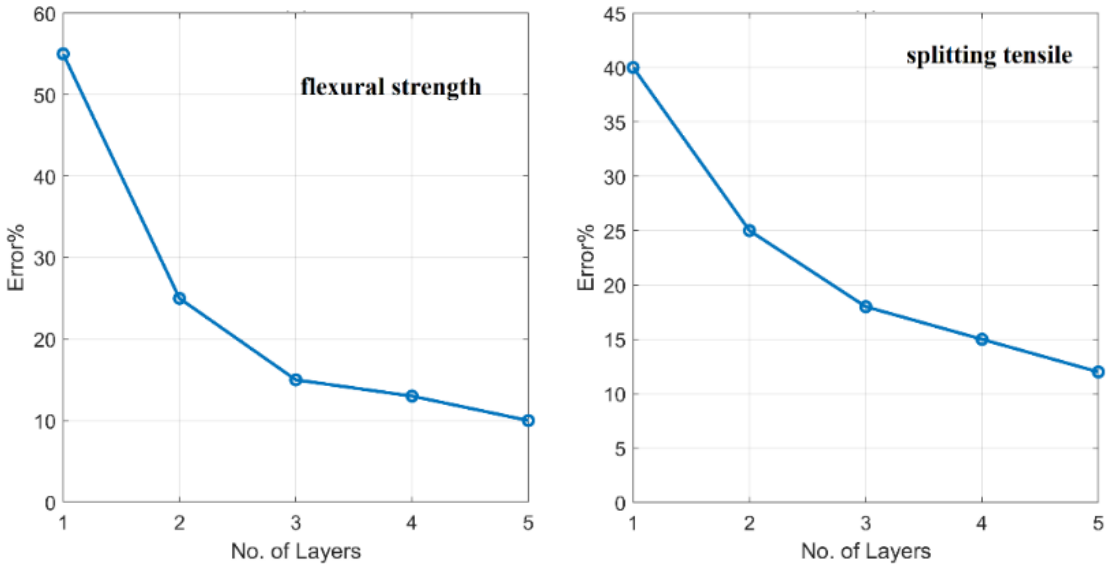


Figure 26. Error (%) reduction in predicting Ff and Ft

The final optimized models yielded strong performance, achieving average prediction accuracies of 96% and 97% for Ff and Ft, respectively, as in Figs. 27 and 28. These results confirm that the DT model is a highly interpretable and moderately high-performing tool, suitable for practical applications where model explainability is essential, such as field-based strength estimation and sustainable mix design optimization.

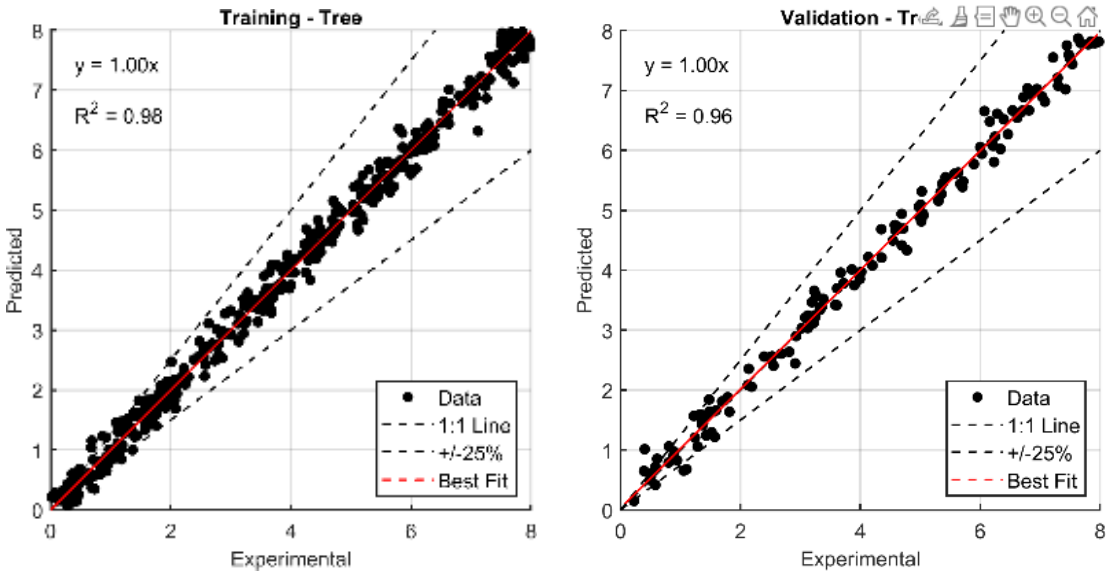


Figure 27. Predicted versus experimental values for Ff using the optimized Decision Tree model

https://doi.org/10.31436/iiumej.v26i3.3730

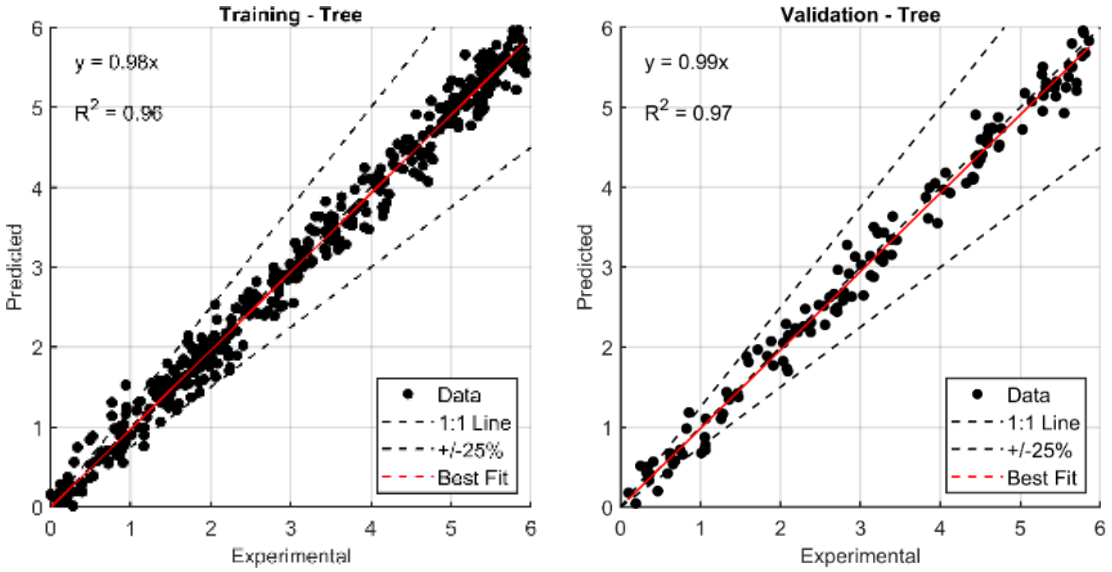


Figure 28. Predicted versus experimental values for Ft using the optimized Decision Tree model

3.1.7. Random Forest (RF) model

The Random Forest (RF) models were developed to predict the flexural and splitting tensile strength of SCGC. These models were generated through a progressive configuration process starting with a single tree of depth one and expanding to three trees with three depth levels. This approach allowed the ensemble to grow in complexity while enabling evaluation of performance improvements across multiple configurations.

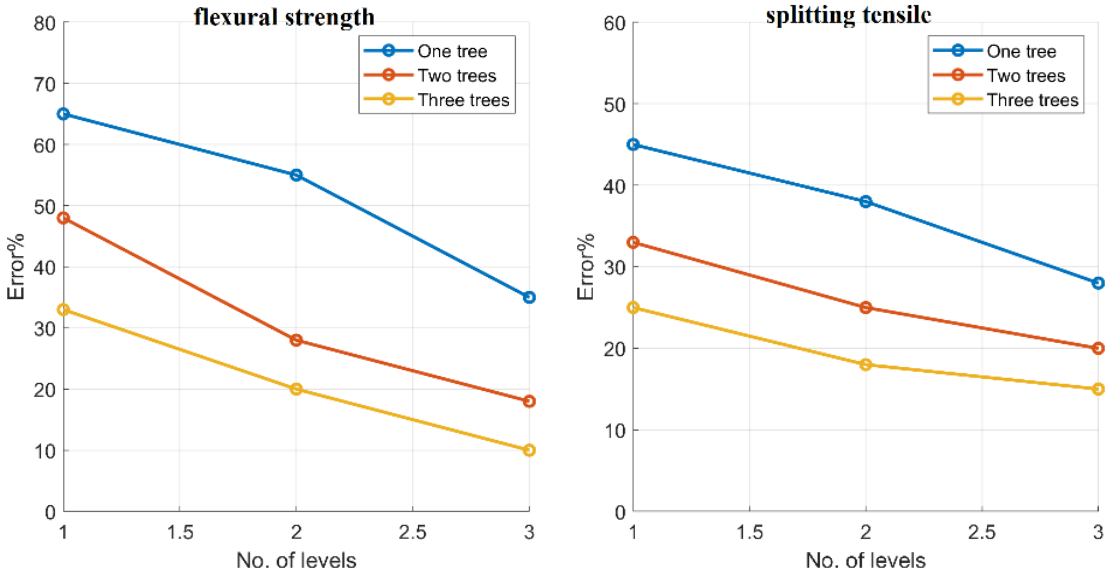


Figure 29. Error (%) reduction trends for Ff and Ft

As depicted in Figure 29, increasing the number of trees and their depth led to a noticeable reduction in prediction error (%) for all target properties. The most substantial improvements occurred in the lower range (from 1 to 2 layers), with diminishing error rates continuing to the third tree level. These results identified the optimal model configuration as three trees with three depth levels, offering the best balance between accuracy and computational efficiency.

https://doi.org/10.31436/iiumej.v26i3.3730

Figure 30 and Figure 31 display the regression results, illustrating how the predicted values align with the experimental data in the training and validation sets for all evaluated strength properties. For Ff, the model achieved (R^2 training = 0.95 and R^2 validation = 0.93), with regression slopes very close to unity ($y = 1.02x$ and $y = 0.98x$), and for Ft, where R^2 ranged from 0.91 to 0.94, confirming the model's capacity to capture the underlying trends without excessive overfitting.

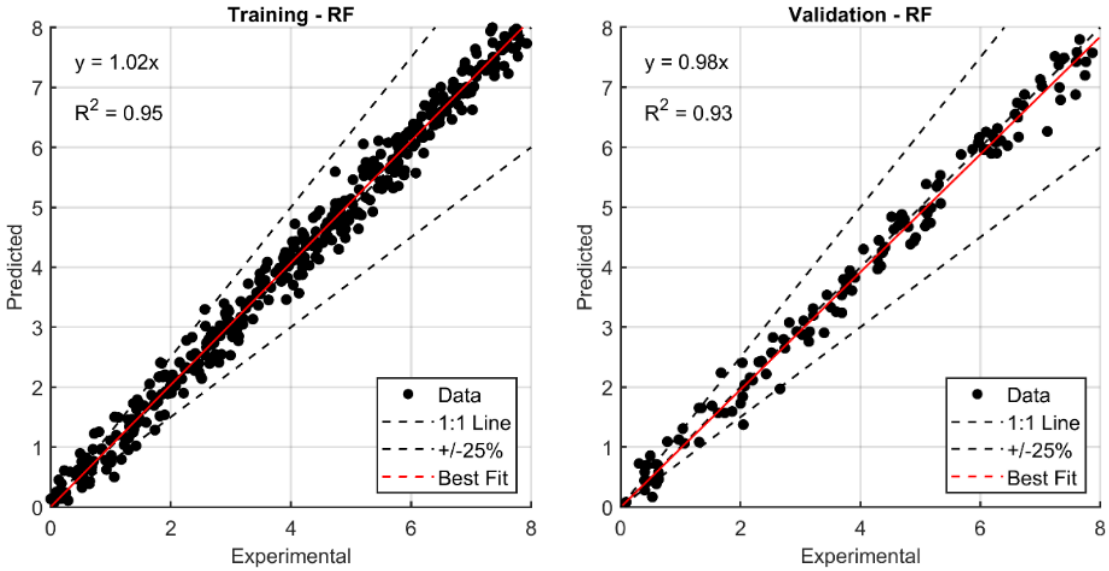


Figure 30. Regression plots for Ff, comparing predicted and experimental values using the optimized Random Forest model

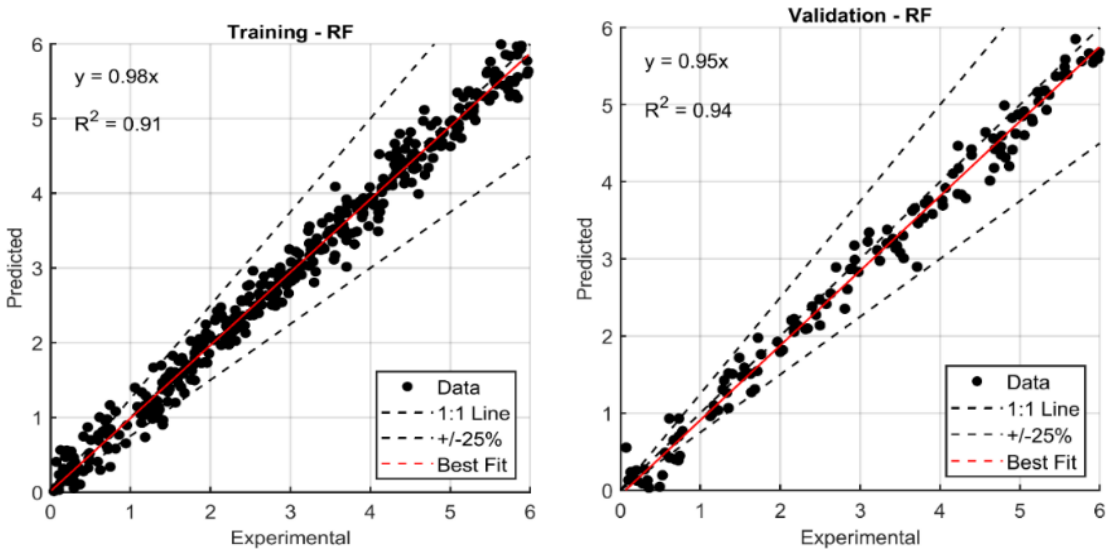


Figure 31. Regression plots for Ft comparing predicted and experimental values using the optimized Random Forest model

Although slightly less accurate than K-NN and SVM models, the RF model offers the advantage of ensemble averaging, which mitigates overfitting and improves robustness against noisy data. Furthermore, RF allows feature importance extraction, making it a valuable tool for identifying the most influential parameters in SCGC mix design optimization.

The SVM and K-NN models emerged as the most accurate and consistent approaches for predicting Ff and Ft. Both models achieved outstanding results, with R^2 values of 0.99 and

RMSE as low as 0.11 MPa, and prediction accuracy reaching 96% on the validation datasets. Figure 32 presents the Ff model performance, where SVM and K-NN clearly outperformed others: the Naïve Bayes (NB) model, by contrast, showed the weakest results with R^2 of 0.76, RMSE of 0.76 MPa, and only 74% accuracy. Similarly, Figure 33 shows the evaluation for splitting tensile strength (Ft), where SVM and K-NN again led the ranking with R^2 values of 0.99, RMSE values of 0.10 MPa, and validation accuracies of 96% and 95%, respectively. While CN2 also provided competitive performance with Ff and Ft prediction accuracies of 95% and 91%, and DT models offered R^2 values of 0.96 and 0.97 for Ff and Ft, respectively, SVM and K-NN offered the best combination of predictive power, generalization, and low error metrics. These findings affirm their suitability for advanced modeling of SCGC mechanical behavior and support their use in sustainable concrete design optimization.

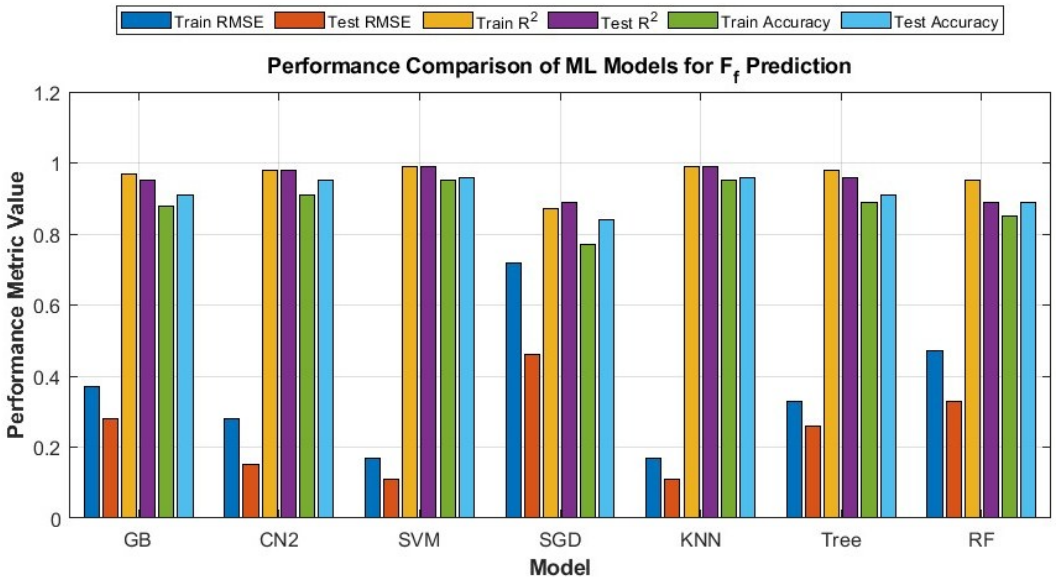


Figure 32. Comparison of machine learning model performance in predicting Ff of SCGC

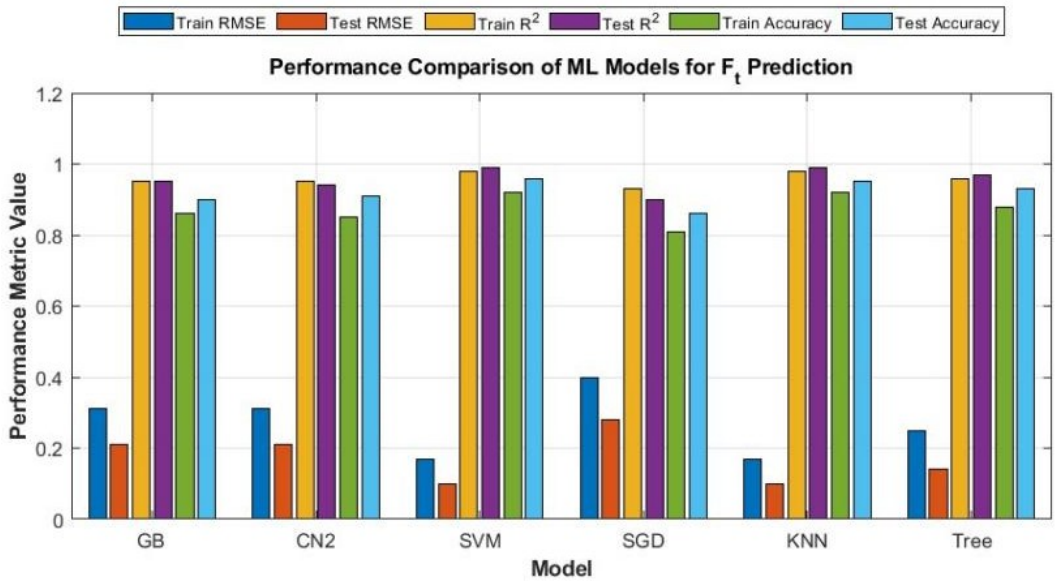


Figure 33. Performance evaluation of developed machine learning models for predicting splitting tensile strength (F_t) of SCGC

https://doi.org/10.31436/iiumej.v26i3.3730

3.1.8. Learning Curve Analysis for Model Generalization

To strengthen the reliability of the reported predictive performance and address the reviewer's request, learning curve analyses were conducted for both Ff and Ft prediction tasks. These analyses assess how the model's predictive accuracy evolves with increasing training data size, using a fixed 75/25 data split. The models evaluated were SVM with KNN. At each analysis step, the training set was incrementally increased from approximately 30 to 408 samples. Both models were retrained from scratch at each training size, and the coefficient of determination (R^2) was calculated on the fixed test set to measure generalization performance.

As shown in Figure 34, the SVM model for Ff prediction achieved high accuracy quickly, reaching an R^2 above 0.99 with only around 100 training samples. Its performance remained consistent across all training sizes, indicating strong generalization. Meanwhile, the KNN model exhibited a more gradual improvement, with R^2 increasing from approximately 0.85 to 0.93 as the training size expanded. This reflects KNN's dependency on a sufficiently large dataset to optimize local predictions.

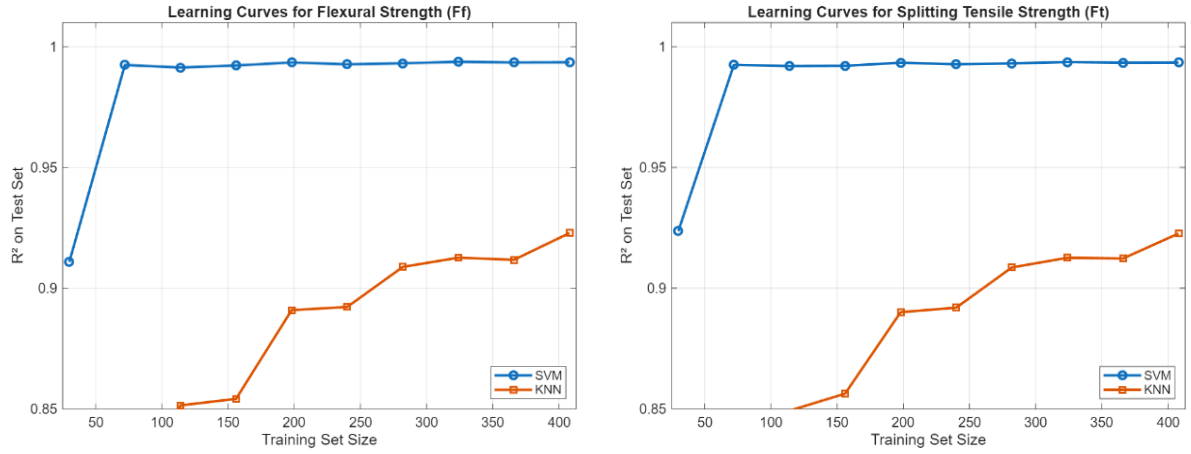


Figure 34. Learning curves of SVM and KNN models for predicting Ff and Ft

A similar trend was observed in the splitting tensile strength (Ft) prediction. The SVM model again reached a stable R^2 of nearly 0.99 early on, while the KNN model improved progressively with additional data, ultimately achieving a comparable R^2 to that of the Ff case. The consistent growth of both curves and the absence of abrupt fluctuations confirm that both models benefit from increased data volume and maintain good generalization without signs of overfitting.

These results validate the earlier reported model performance and demonstrate that the high R^2 values (approaching 0.99) are reproducible under a held-out test condition. The learning curve analysis provides clear visual evidence of the robustness of the SVM model and the reliability of KNN when supported by sufficient training data.

3.2. Response Surface Methodology (RSM) Analysis, Diagnostic Validation, and Hoffman and Gardener Sensitivity Analysis

To complement machine learning techniques with interpretable mathematical modeling, RSM was employed to estimate Ff and Ft of SCGC. RSM enables the derivation of second-order polynomial equations that describe the relationships between key input parameters GGBS, FA, NaOH, Na_2SiO_3, and curing age and the target strength properties. The approach also provides a suite of statistical diagnostics and visualization tools that validate model quality and generalizability.

https://doi.org/10.31436/iiumej.v26i3.3730

For Ff, the RSM model achieved an R^2 of 0.9879, with an adjusted R^2 of 0.9796 and a predicted R^2 of 0.9727. The minimal difference between adjusted and predicted values (0.0069) confirms strong internal consistency. The model's Adequate Precision was calculated at 85.318, significantly higher than the minimum threshold of 4, indicating excellent signal strength. Similarly, the RSM model for Ft showed outstanding performance, with an R^2 of 0.9878, adjusted R^2 of 0.9863, and predicted R^2 of 0.9801, resulting in a minimal difference of 0.0062, well within acceptable bounds. Although the Adequate Precision for Ft was not shown graphically, it was confirmed in the regression output to exceed the minimum value, reinforcing model robustness.

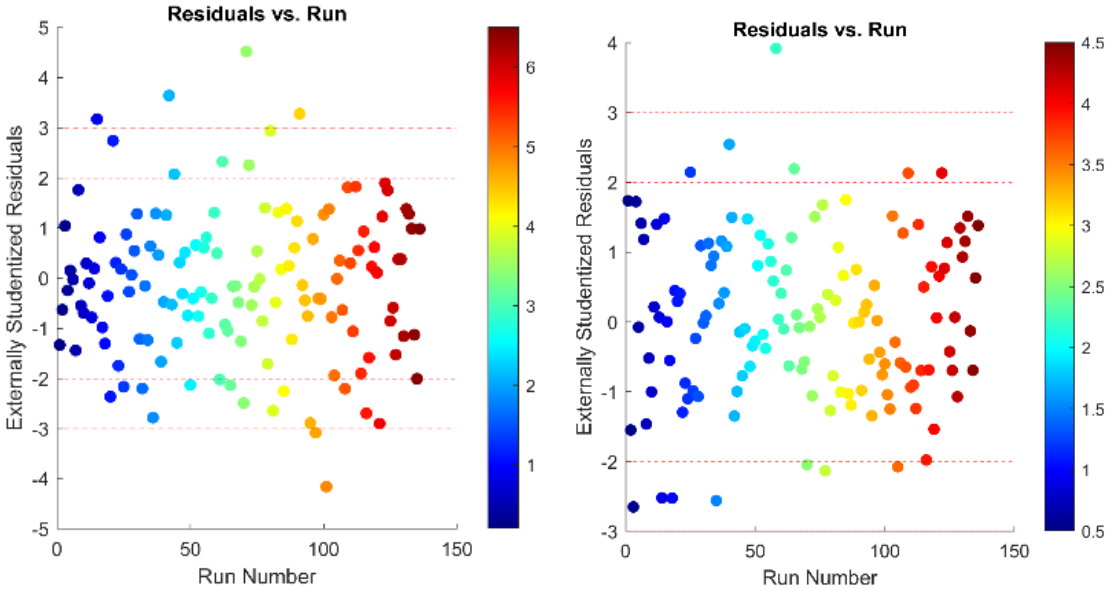


Figure 35. Residuals versus experimental run order for Ff and Ft

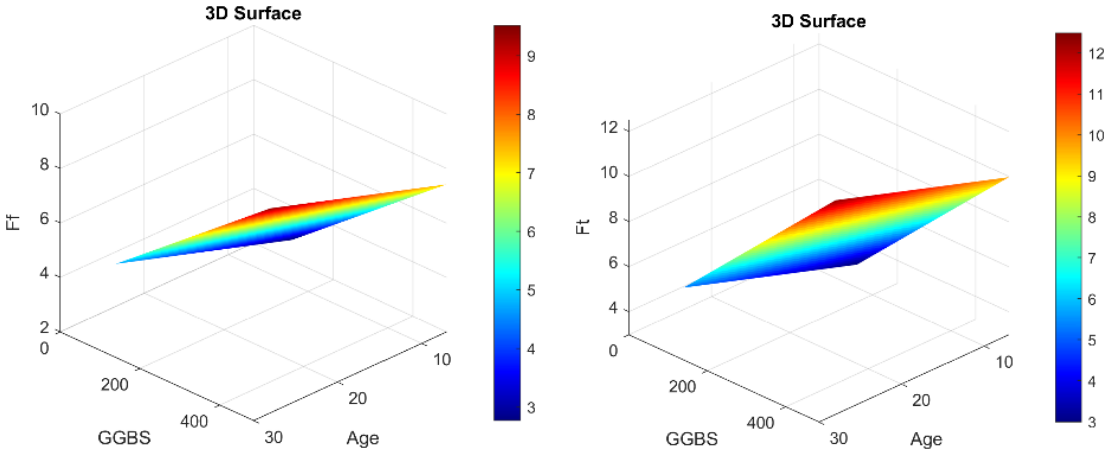


Figure 36. 3D response surface showing the interaction between GGBS and Age on Ff and Ft

The residuals vs. run number plots, Figure 35, show randomly scattered externally studentized residuals, indicating that residuals are independent, homoscedastic, and free from run-order bias in both models. This randomness supports the assumption that model errors are normally distributed and independent of the experimental order. The 3D surface plot of GGBS and Age (Figure 36). illustrates a clear rising plane, confirming their joint positive influence on Ft.

https://doi.org/10.31436/iiumej.v26i3.3730

Regarding influence diagnostics, the Cook's Distance plots for Ff and Ft, as in Figure 37, demonstrate that all data points fall well below the critical threshold. This indicates that no individual observation has a disproportionately large impact on the regression output in either model. The uniformity across these plots affirms the structural stability and fairness of the models.

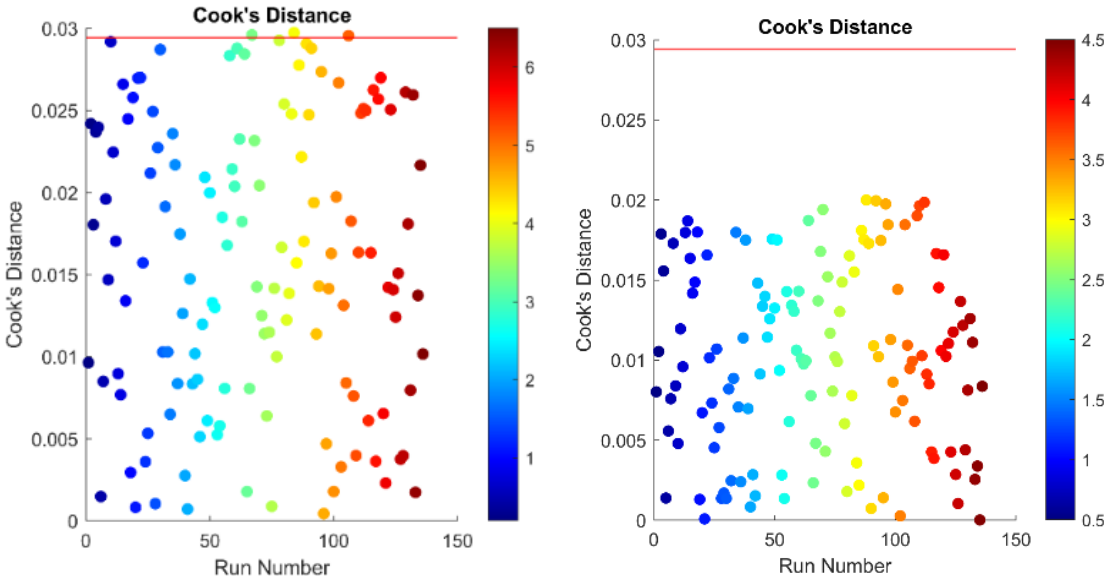


Figure 37. Cook's Distance plot for Ff, Ft model

The predicted vs. actual plots in Figure 38 for both strength models exhibit a tight clustering of points along the ideal 1:1 reference line. This indicates a strong correlation between experimental and predicted values and confirms the models' high predictive capability. These plots show the model's accuracy in replicating real-world outcomes.

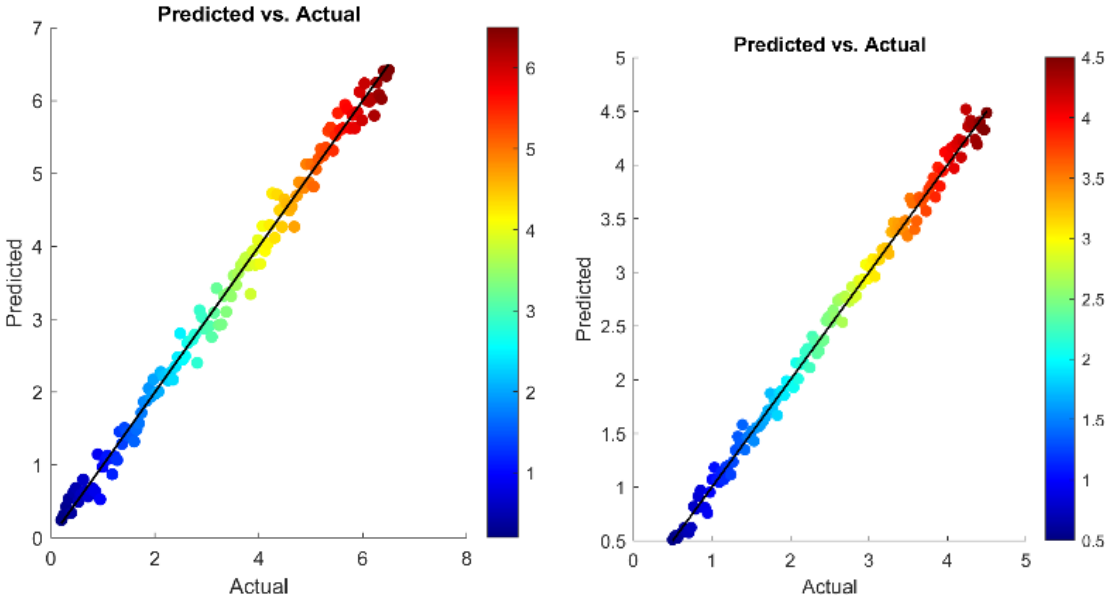


Figure 38. Predicted versus actual flexural strength and tensile strength values

Leverage diagnostics also support model validity. In both the Ff model and the Ft model, as shown in Figure 39, all leverage values remain well within the safe zone, confirming that no single observation has an excessive statistical influence on the fitted equation. The distribution of leverage values suggests that the experimental design is well-balanced.

https://doi.org/10.31436/iiumej.v26i3.3730

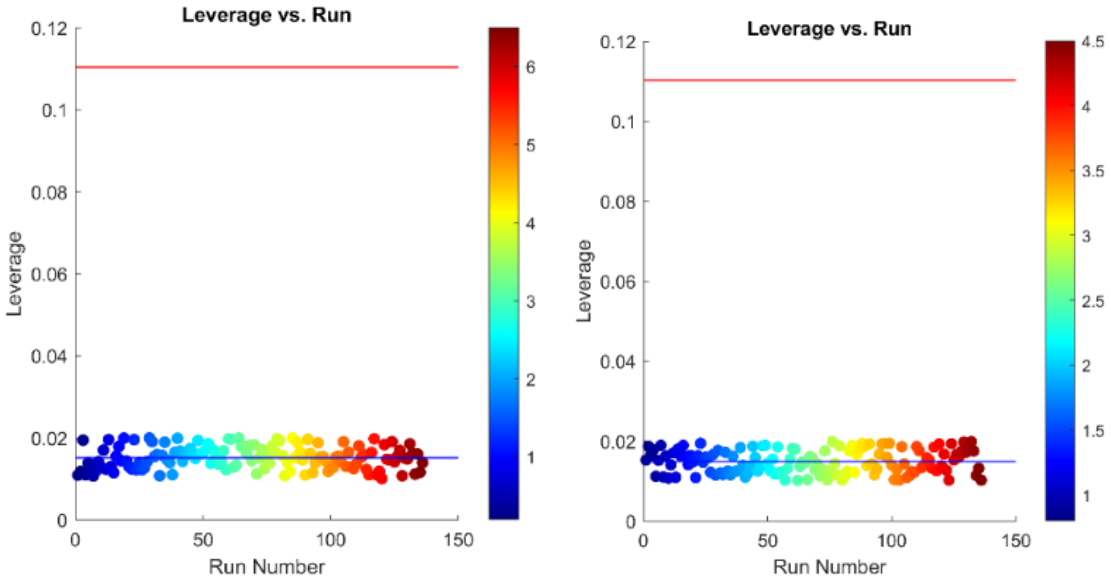


Figure 39. Leverage values plotted against run number

Lastly, the DFBETAS plots for the intercept term, Figure 40, show that the sensitivity of the regression intercept to any specific observation is minimal. All data points fall within the ±0.2 threshold, ensuring that the model coefficients remain stable and unaffected by isolated experimental runs.

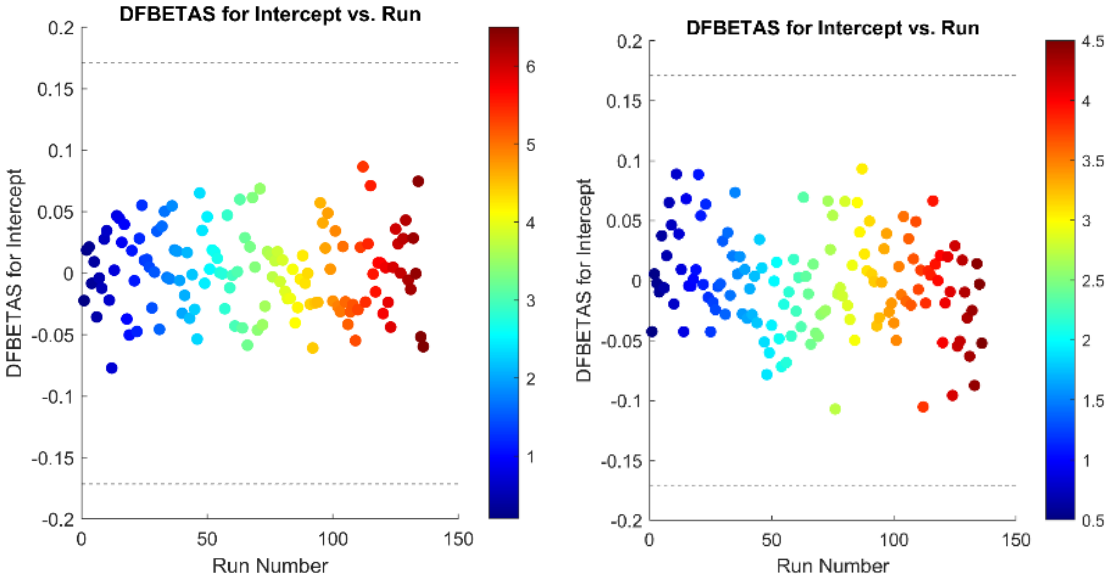


Figure 40. DFBETAS for the intercepting term plotted versus run number

These combined diagnostics confirm that the RSM models for Ff and Ft are statistically sound, structurally stable, and capable of effectively generalizing to unseen data. The consistently high R^2 values, low error distributions, and absence of influential outliers support the conclusion that RSM provides a transparent and reliable framework for modeling the mechanical performance of geopolymer concrete based on compositional and curing parameters.

The equations derived from RSM, expressed using actual factor values, provide practical tools for predicting strength outcomes in SCGC. These closed-form models account for main

https://doi.org/10.31436/iiumej.v26i3.3730

effects, interaction effects, and nonlinear contributions of critical mix parameters presented in Eq. (18) and Eq. (19).

$$\text{Flexural Strength (Ff)} = 5.62342 + 0.014130\,\text{GGBS} + 1.000000\,\text{FA} + 0.016490\,\text{NaOH} + 1.000000\,\text{Na2SiO3} + 0.034523\,\text{Age} + 0.000010\,(\text{GGBS}\cdot\text{FA}) + 0.000060\,(\text{GGBS}\cdot\text{NaOH}) + 1.000000\,(\text{GGBS}\cdot\text{Na2SiO3}) + 0.000024\,(\text{GGBS}\cdot\text{Age}) + 1.000000\,(\text{FA}\cdot\text{NaOH}) + 1.000000\,(\text{FA}\cdot\text{Na2SiO3}) + 1.000000\,(\text{FA}\cdot\text{Age}) + 0.000912\,(\text{NaOH}\cdot\text{Na2SiO3}) + 0.000019\,(\text{NaOH}\cdot\text{Age}) + 1.000000\,(\text{Na2SiO3}\cdot\text{Age}) + (\text{GGBS})^2 + (\text{FA})^2 + (\text{NaOH})^2 + (\text{Na2SiO3})^3 + (\text{Age})^2 \quad (18)$$

$$\text{Tensile strength Ft} = 4.86932 + 0.009817\,\text{GGBS} + 1.000000\,\text{FA} + 0.016063\,\text{NaOH} + 1.000000\,\text{Na2SiO3} + 0.017814\,\text{Age} + 3.07324 \times 10 - 6\,(\text{GGBS}\cdot\text{FA}) + 0.000050\,(\text{GGBS}\cdot\text{NaOH}) + 1.000000\,(\text{GGBS}\cdot\text{Na2SiO3}) + 0.000037\,(\text{GGBS}\cdot\text{Age}) + 1.000000\,(\text{FA}\cdot\text{NaOH}) + 1.000000\,(\text{FA}\cdot\text{Na2SiO3}) + 1.000000\,(\text{FA}\cdot\text{Age}) + 0.000847\,(\text{NaOH}\cdot\text{Na2SiO3}) + 0.000026\,(\text{NaOH}\cdot\text{Age}) + 1.000000\,(\text{Na2SiO3}\cdot\text{Age}) + (\text{GGBS})^2 + (\text{FA})^2 + (\text{NaOH})^2 + (\text{Na2SiO3})^3 + 0.000468\,(\text{Age})^2 \quad (19)$$

Figure 41 illustrates the Hoffman analysis conducted for the two primary strength parameters, Ff and Ft. The study provides a quantitative breakdown of the contribution of each input factor, FA, GGBS, NaOH, Na_2SiO_3, and curing Age, to the variation observed in the model outputs.

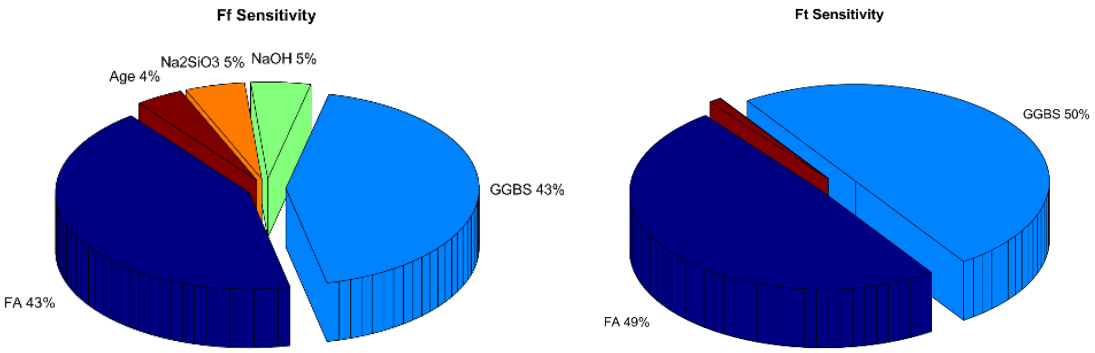


Figure 41. Hoffman and Gardner's sensitivity analysis

Binder constituents had the most significant influence on flexural and tensile strength development in SCGC mixtures. Specifically, fly ash and GGBS displayed equally dominant effects on flexural strength, highlighting their critical role in load-bearing resistance. In contrast, alkaline activators such as sodium hydroxide and sodium silicate contributed only marginally, with curing age also exerting a minor impact. When examining tensile strength, the dominance of binder materials became even more pronounced, with fly ash and GGBS accounting for nearly the entire influence. At the same time, the remaining variables showed virtually no contribution. This trend illustrates that achieving optimal strength outcomes in geopolymer concrete relies heavily on the proportioning and quality of binder materials. The negligible impact of other mix parameters suggests that efforts to refine strength performance can be streamlined by prioritizing binder optimization, reducing complexity in the design process, and offering more consistent control over mechanical outcomes in sustainable concrete applications.

https://doi.org/10.31436/iiumej.v26i3.3730

4. LIMITATION OF THIS STUDY

One major constraint lies in the dataset size and composition. Although 544 experimental records were compiled from reliable sources, the data originated from different laboratories with varying testing protocols, material purities, and environmental conditions. Such variability may introduce hidden biases and affect model training and validation consistency. Moreover, the study focused solely on predicting two mechanical properties, Ff and Ft, without addressing other essential performance metrics such as modulus of elasticity, shrinkage, creep, or long-term durability indicators. These properties are equally important in practical applications and warrant inclusion in future research to establish a more holistic prediction framework.

Additionally, while ML models like K-NN and SVM offered excellent predictive accuracy, they remain limited in interpretability, especially for stakeholders unfamiliar with data science. The "black-box" nature of these models can make extracting clear decision rules for field applications challenging. Although Response Surface Methodology (RSM) improved model transparency, its predictive capability is confined to the parameter range used during model fitting, making extrapolation beyond the tested domain unreliable. Furthermore, the sensitivity analysis adopted in this study primarily captured linear relationships, which may overlook complex, higher-order interactions among input variables. To strengthen the applicability of such models in real-world construction scenarios, future work should integrate larger, more diverse datasets, explore broader mechanical and durability parameters, and adopt advanced interpretability tools or global sensitivity analysis techniques.

5. CONCLUSION

This study provides a robust and comprehensive framework for predicting the flexural strength (Ff) and splitting tensile strength (Ft) of Self-Compacting Geopolymer Concrete (SCGC) using a dual approach that integrates machine learning (ML) techniques and Response Surface Methodology (RSM). Addressing a notable gap in the current literature, where compressive strength has been extensively studied while other mechanical properties like Ff and Ft remain underexplored, this research contributes to the growing domain of data-driven concrete modeling aimed at sustainable construction solutions.

A diverse suite of eight machine learning algorithms, including K-Nearest Neighbors (K-NN), Support Vector Machines (SVM), CN2 rule induction, Random Forest (RF), Gradient Boosting (GB), Decision Trees (DT), Naïve Bayes (NB), and Stochastic Gradient Descent (SGD), was evaluated. Among these, K-NN and SVM exhibited the highest predictive accuracy, with R^2 values reaching 0.99 and root mean squared errors (RMSE) as low as 0.10 for both Ff and Ft. These results confirm their suitability for modeling the nonlinear and multivariate relationships characteristic of geopolymer concrete mix behavior. While CN2, DT, GB, and RF models also showed strong performance ($R^2 \geq 0.93$), the Naïve Bayes model underperformed, underscoring the need for algorithms to accommodate complex feature interdependencies.

Complementing the machine learning analysis, RSM provided interpretable polynomial equations capable of estimating strength outcomes based on actual mix parameters. These models achieved high statistical validity, with R^2 values exceeding 0.98 and strong agreement between adjusted and predicted R^2 scores, indicating excellent model generalization. The models passed rigorous diagnostics, including leverage, Cook's Distance, and DFBETAS, confirming the structural integrity of the regression outputs.

Furthermore, Hoffman and Gardner's sensitivity analysis revealed that binder materials, Fly Ash (FA) and Ground Granulated Blast Furnace Slag (GGBS), are the most influential factors for Ff and Ft, collectively contributing over 90% to strength variation. In contrast, sodium-based activators (NaOH, Na_2SiO_3) and curing age had minimal impact, simplifying the optimization process by focusing design efforts on binder proportions.

Finally, integrating advanced ML algorithms with interpretable RSM models offers a powerful toolkit for predicting and optimizing SCGC mechanical performance. This hybrid framework enhances the accuracy and reliability of strength estimation, improving practical decision-making in mix design for sustainable concrete technologies. Future studies may build upon this foundation by incorporating additional performance parameters such as durability, shrinkage, and creep, as well as deploying these models within digital twin platforms for real-time monitoring and adaptive construction management.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support provided by Universitas Muhammadiyah Yogyakarta (UMY).

REFERENCES

[1] Heshmati, M., Sheikh, M. N., & Hadi, M. N. S. (2025). A comprehensive review of the fresh and hardened characteristics of self-compacting geopolymer concrete. Journal of Sustainable Cement-Based Materials, 14(4), 816–835. https://doi.org/10.1080/21650373.2025.2461182

[2] Gopalakrishna, B., & Pasla, D. (2024). Study of Ambient Cured Fly Ash-GGBS-Metakaolin-Based Geopolymers Mortar. Lecture Notes in Civil Engineering, 440. https://doi.org/10.1007/978-981-99-7464-1_3

[3] Shahrokh, S., Armaghan, S., & Danandeh, A. M. (2025). Service Life Prediction and Life Cycle Assessment of Pozzolanic Concretes. 54(4), 29–36. 10.22034/ceej.2024.61027.2338

[4] Inqiad, W. Bin, Siddique, M. S., Alarifi, S. S., Butt, M. J., Najeh, T., & Gamil, Y. (2023). Comparative analysis of various machine learning algorithms to predict 28-day compressive strength of Self-compacting concrete. Heliyon, 9(11). https://doi.org/10.1016/j.heliyon.2023.e22036

[5] Sukontasukkul, P., Intarabut, D., Phoo-ngernkham, T., Suksiripattanapong, C., Zhang, H., & Chindaprasirt, P. (2023). Self-compacting steel fibers reinforced geopolymer: Study on mechanical properties and durability against acid and chloride attacks. Case Studies in Construction Materials, 19. https://doi.org/10.1016/j.cscm.2023.e02298

[6] Sun, C., Wang, K., Liu, Q., Wang, P., & Pan, F. (2023). Machine-Learning-Based Comprehensive Properties Prediction and Mixture Design Optimization of Ultra-High-Performance Concrete. Sustainability (Switzerland), 15(21). https://doi.org/10.3390/su152115338

[7] Ahmad, A., Ahmad, W., Chaiyasarn, K., Ostrowski, K. A., Aslam, F., Zajdel, P., & Joyklad, P. (2021). Prediction of geopolymer concrete compressive strength using novel machine learning algorithms. Polymers, 13(19). https://doi.org/10.3390/polym13193389

[8] Saha, P., Sapkota, S. C., Das, S., & Kwatra, N. (2024). Prediction of fresh and hardened properties of self-compacting concrete using ensemble soft learning techniques. Multiscale and Multidisciplinary Modeling Experiments and Design, 7(5), 4923–4945. https://doi.org/10.1007/s41939-024-00423-5

[9] de-Prado-Gil, J., Palencia, C., Silva-Monteiro, N., & Martínez-García, R. (2022). To predict the compressive strength of self compacting concrete with recycled aggregates utilizing ensemble machine learning models. Case Studies in Construction Materials, 16. https://doi.org/10.1016/j.cscm.2022.e01046

[10] Biswas, R., Kumar, M., Singh, R. K., Alzara, M., El Sayed, S. B. A., Abdelmongy, M., Yosri, A. M., & Yousef, S. E. A. S. (2023). A novel integrated approach of RUNge Kutta optimizer and ANN for estimating compressive strength of self-compacting concrete. Case Studies in Construction Materials, 18. https://doi.org/10.1016/j.cscm.2023.e02163

[11] Dong, B., Zhao, S., Zhang, Y., Zhang, Y., Yu, Y., & Yang, J. (2025). Physics-based probabilistic analysis of corrosion initiation in alkali-activated slag concrete assisted by machine learning. Construction and Building Materials, 471. https://doi.org/10.1016/j.conbuildmat.2025.140661

[12] Iqbal, M., Nazar, S., Yang, J., & Mahmoud, H. A. (2025). Integrated machine learning and response surface methodology for comprehensive rheological characterization of low-carbon binders. Case Studies in Construction Materials, 22. https://doi.org/10.1016/j.cscm.2025.e04475

[13] Parhi, S. K., Dwibedy, S., & Panigrahi, S. K. (2024). AI-driven critical parameter optimization of sustainable self-compacting geopolymer concrete. Journal of Building Engineering, 86. https://doi.org/10.1016/j.jobe.2024.108923

[14] Aneja, S., Sharma, A., Gupta, R., & Yoo, D. Y. (2021). Bayesian regularized artificial neural network model to predict strength characteristics of fly-ash and bottom-ash based geopolymer concrete. Materials, 14(7). https://doi.org/10.3390/ma14071729

[15] Jiang, H., Liu, G., Alyami, H., Alharbi, A., Jameel, M., & Khadimallah, M. A. (2022). Intelligence decision mechanism for prediction of compressive strength of self-compaction green concrete via neural network. Journal of Cleaner Production, 340. https://doi.org/10.1016/j.jclepro.2022.130580

[16] Mazumder, E. A., & Prasad Meesaraganda, L. V. (2023). Probabilistic Estimation for Mechanical Properties of Self-Compacting Geopolymer Concrete Using Machine Learning Technique. Arabian Journal for Science and Engineering, 48(10). https://doi.org/10.1007/s13369-023-07866-x

[17] Aliabdo, A. A., Abd Elmoaty, A. E. M., & Salem, H. A. (2016). Effect of water addition, plasticizer and alkaline solution constitution on fly ash based geopolymer concrete performance. Construction and Building Materials, 121, 694–703. https://doi.org/10.1016/J.CONBUILDMAT.2016.06.062

[18] Phoo-Ngernkham, T., Phiangphimai, C., Damrongwiriyanupap, N., Hanjitsuwan, S., Thumrongvut, J., & Chindaprasirt, P. (2018). A Mix Design Procedure for Alkali-Activated High-Calcium Fly Ash Concrete Cured at Ambient Temperature. Advances in Materials Science and Engineering, 2018. https://doi.org/10.1155/2018/2460403

[19] Arun, B. R., Srishaila, J. M., S, M. K., Algur, V., K, K., & Tanu, H. (2024). Prediction of machine learning application in the development of novel sustainable self-compacting geopolymer concrete. Research on Engineering Structures and Materials, x(xxxx), 1–22. https://doi.org/10.17515/resm2024.354ma0716rs

[20] Lavanya, G., & Jegan, J. (2015). Durability Study on High Calcium Fly Ash Based Geopolymer Concrete. Advances in Materials Science and Engineering, 2015. https://doi.org/10.1155/2015/731056

[21] Deilami, S., Aslani, F., & Elchalakani, M. (2019). An experimental study on the durability and strength of SCC incorporating FA, GGBS and MS. Proceedings of the Institution of Civil Engineers: Structures and Buildings, 172(5). https://doi.org/10.1680/jstbu.17.00129

[22] Rangan, B., & Hardjito, D. (2005). Studies on fly ash-based geopolymer concrete. Proc. 4th World …, November.

[23] Nishanth, L., & Patil, D. N. N. (2022). Experimental evaluation on workability and strength characteristics of self-consolidating geopolymer concrete based on GGBFS, flyash and alccofine. Materials Today: Proceedings, 59. https://doi.org/10.1016/j.matpr.2021.10.200

[24] Vijaya Prasad, B., Paul Daniel, A. P., Anand, N., & Yadav, S. K. (2022). Strength and microstructure behaviour of high calcium fly ash based sustainable geo polymer concrete.

Journal of Engineering, Design and Technology, 20(2). https://doi.org/10.1108/JEDT-03-2021-0178

[25] Alam, M. F., Shubham, K., Kumar, S., & Srivastava, A. K. L. (2024). Enhancing high-strength self-compacting concrete properties through Nano-silica: analysis and prediction of mechanical strengths. Journal of Building Pathology and Rehabilitation, 9(1). https://doi.org/10.1007/s41024-024-00386-7

[26] Tho-In, T., Sata, V., Chindaprasirt, P., & Jaturapitakkul, C. (2012). Pervious high-calcium fly ash geopolymer concrete. Construction and Building Materials, 30. https://doi.org/10.1016/j.conbuildmat.2011.12.028

[27] Bheel, N., Awoyera, P., Tafsirojjaman, T., Hamah Sor, N., & sohu, S. (2021). Synergic effect of metakaolin and groundnut shell ash on the behavior of fly ash-based self-compacting geopolymer concrete. Construction and Building Materials, 311. https://doi.org/10.1016/j.conbuildmat.2021.125327

[28] Pirhadi, N., Tang, X., Yang, Q., & Kang, F. (2019). A new equation to evaluate liquefaction triggering using the response surface method and parametric sensitivity analysis. Sustainability (Switzerland), 11(1). https://doi.org/10.3390/su11010112

[29] Fontanari, T., Fróes, T. C., & Recamonde-Mendoza, M. (2022). Cross-validation Strategies for Balanced and Imbalanced Datasets. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 13653 LNAI. https://doi.org/10.1007/978-3-031-21686-2_43

[30] Li, Z., Pei, T., Ying, W., Srubar III, W. V., Zhang, R., Yoon, J., Ye, H., Dabo, I., & Radlińska, A. (2024). Simulation-Based Transfer Learning for Concrete Strength Prediction. In Rilem Bookseries (Vol. 48). https://doi.org/10.1007/978-3-031-53389-1_98

[31] Shaikhina, T., & Khovanova, N. A. (2017). Handling limited datasets with neural networks in medical applications: A small-data approach. Artificial Intelligence in Medicine, 75, 51–63. https://doi.org/10.1016/j.artmed.2016.12.003

[32] Arthi, K., Sankaradass, V., Parveen, N., & Muralidharan, J. (2023). Methods of cross-validation and bootstrapping. In Toward Artificial General Intelligence: Deep Learning, Neural Networks, Generative AI. https://doi.org/10.1515/9783111323749-007

[33] Hong, J.-S., Lee, J., & Sim, M. K. (2024). Concise rule induction algorithm based on one-sided maximum decision tree approach. Expert Systems with Applications, 237. https://doi.org/10.1016/j.eswa.2023.121365

[34] Wickramasinghe, I., & Kalutarage, H. (2021). Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. Soft Computing, 25(3), 2277–2293. https://doi.org/10.1007/s00500-020-05297-6

[35] Jin, Z., Hei, T. G., & Jin, K. (2024). Application of support vector machines for categorizing biological and medical data. Proceedings of SPIE - The International Society for Optical Engineering, 13105. https://doi.org/10.1117/12.3026871

[36] Saha, A., Yadav, R. K., & Dwivedi, R. K. (2024). Efficient Convergence Analysis of Stochastic Gradient Descent for Large-Scale Optimization. 2024 International Conference on Optimization Computing and Wireless Communication, ICOCWC 2024. https://doi.org/10.1109/ICOCWC60930.2024.10470675

[37] Liang, C., Xia, S., & Chen, Z. (2019). Improvement k-Nearest Neighbor Classification Algorithm Based on Reference Points. Jisuanji Gongcheng/Computer Engineering, 45(2). https://doi.org/10.19678/j.issn.1000-3428.0049901

[38] Rutkowski, L., Jaworski, M., & Duda, P. (2020). Decision Trees in Data Stream Mining. In Studies in Big Data (Vol. 56). https://doi.org/10.1007/978-3-030-13962-9_3

[39] Tran, T. T., Phan, N. Q., & Huynh, H. X. (2025). Random Forest Model Parameters Optimization. In Communications in Computer and Information Science: Vol. 2191 CCIS. https://doi.org/10.1007/978-981-97-9616-8_19

[40] Podder, S., & Mukherjee, S. (2024). RESPONSE SURFACE METHODOLOGY (RSM) AS A TOOL IN PHARMACEUTICAL FORMULATION DEVELOPMENT. Asian Journal of Pharmaceutical and Clinical Research, 17(11), 18–25. https://doi.org/10.22159/ajpcr.2024v17i11.52149

[41] Mahmoodi, V., & Sargolzaei, J. (2014). Optimization of photocatalytic degradation of naphthalene using nano-TiO2/UV system: statistical analysis by a response surface methodology. Desalination and Water Treatment, 52(34–36). https://doi.org/10.1080/19443994.2013.861774

[42] Onyelowe, K. C., Kamchoom, V., Hanandeh, S., Ebid, A. M., Llamuca Llamuca, J. L., Cayán Martínez, J. C., Rose, E., Awoyera, P., & Avudaiappan, S. (2025). Predicting the strengths of basalt fiber reinforced concrete mixed with fly ash using AML and Hoffman and Gardener techniques. Scientific Reports, 15(1). https://doi.org/10.1038/s41598-025-96420-w

[43] Paleari, L., Movedi, E., Zoli, M., Burato, A., Cecconi, I., Errahouly, J., Pecollo, E., Sorvillo, C., & Confalonieri, R. (2021). Sensitivity analysis using Morris: Just screening or an effective ranking method? Ecological Modelling, 455. https://doi.org/10.1016/j.ecolmodel.2021.109648

[44] Hoffman, F. O., & Gardner, R. H. (1983). Evaluation of uncertainties in radiological assessment models. Radiological Assessment: A Textbook on Environmental Dose Analysis.

[45] Wu, Y., & Liu, S. (2014). A suggestion for computing objective function in model calibration. Ecological Informatics, 24. https://doi.org/10.1016/j.ecoinf.2014.08.002

[46] Robeson, S. M., & Willmott, C. J. (2023). Decomposition of the mean absolute error (MAE) into systematic and unsystematic components. Plos One, 18(2 February). https://doi.org/10.1371/journal.pone.0279774

[47] Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science, 7. https://doi.org/10.7717/PEERJ-CS.623

[48] Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 30(1). https://doi.org/10.3354/cr030079

[49] Turney, S. (2022). Coefficient of determination (R2): Calculation and interpretation. Scribbr.

https://doi.org/10.31436/iiumej.v26i3.3821

INITIAL INVESTIGATION ON IMPROVING THE PHYSICOMECHANICAL PROPERTIES OF CLAYEY SAND USING CLAY BRICK DUST

UMME SAIMA NUR, NORHIDAYU BINTI KASIM*, WAN NUR AIFA WAN AZAHAR

Department of Civil Engineering, International Islamic University Malaysia, Jln Gombak, 53100 Kuala Lumpur, Selangor, Malaysia

**Corresponding author: ayukasim@iium.edu.my*

(Received: 30 June 2025; Accepted: 27 August 2025; Published online: 9 September 2025)

ABSTRACT: Clayey sand is commonly considered unsuitable for construction due to its high compressibility, low shear strength, and susceptibility to erosion and settlement. The research focuses on enhancing the physical and mechanical properties of the soil by incorporating clay brick dust (CBD) at varying proportions of 15, 20, and 25% by weight. A comprehensive laboratory testing was conducted, including particle size distribution, Atterberg limits, moisture content, Standard Proctor compaction, and California Bearing Ratio (CBR). These tests were used to assess the physical and mechanical behaviour of both untreated and treated soils. The results indicate that the inclusion of CBD significantly improves the soil's strength, as reflected by increased CBR values. However, higher percentages of CBD lead to reductions in moisture content and maximum dry density. The study concluded that the optimum percentage of CBD mixture was found to be 20% by weight, offering the best balance between improved strength and acceptable compaction properties. This study concludes that CBD is a viable and sustainable material for stabilizing clayey sand, making it more suitable for geotechnical and construction applications.

ABSTRAK: Pasir berlempung kebiasaannya tidak sesuai bagi tujuan pembinaan disebabkan oleh kemampatannya yang tinggi, kekuatan ricih yang rendah, serta kerentanan terhadap hakisan dan pemendapan. Kajian ini memfokuskan kepada penambahbaikan sifat fizikal dan mekanikal tanah dengan campuran serbuk bata tanah liat (CBD) pada kadar berbeza iaitu 15%, 20% dan 25% mengikut berat. Ujian makmal menyeluruh telah dijalankan merangkumi taburan saiz zarah, had-had Atterberg, kandungan lembapan, pemadatan Proktor piawai, dan ujian Nisbah Galas California (CBR). Ujian-ujian ini digunakan bagi menilai sifat fizikal dan mekanikal tanah sebelum dan selepas rawatan. Dapatan kajian menunjukkan bahawa penambahan CBD secara signifikan meningkatkan kekuatan tanah, seperti yang dibuktikan melalui peningkatan nilai CBR. Walau bagaimanapun, peratusan CBD yang lebih tinggi mengakibatkan pengurangan kandungan lembapan dan ketumpatan kering maksimum. Kajian ini merumuskan bahawa campuran optimum CBD adalah pada 20% berat, yang memberikan keseimbangan terbaik antara peningkatan kekuatan dan sifat pemadatan terbaik. Kajian ini membuktikan bahawa CBD merupakan bahan berdaya guna dan mampan bagi menstabilkan pasir berlempung, sekaligus menjadikannya lebih sesuai untuk aplikasi geoteknik dan pembinaan.

KEY WORDS: *Clayey sand, Soil Stabilization, Clay brick dust, Geotechnical Properties, California Bearing Ratio (CBR), Compaction characteristics*

https://doi.org/10.31436/iiumej.v26i3.3821

1. INTRODUCTION

Clayey sand, a composite of fine clay and coarse sand particles, presents notable challenges in civil and environmental engineering applications. It is characterized by moderate to high compressibility, low shear strength, and a tendency to erode and settle, rendering it less ideal for construction [1]. Its high moisture retention may cause waterlogging, while its uneven particle distribution complicates compaction and shaping during site preparation [2]. Despite these drawbacks, enhancing the geotechnical behavior of clayey sand in terms of strength, compressibility, and moisture response is crucial for its application in infrastructure.

Soil stabilization is a widely used method to enhance soil performance and ensure reliable ground and foundation support. It improves load-bearing capacity and durability to meet safety standards. Various stabilizing agents have been used, including stone dust, calcium chloride, ceramic dust, lime, cement, bitumen, and brick dust, each offering distinct benefits depending on soil conditions [3].

Brick dust, derived from fired clay bricks, has gained attention due to its water resistance, refractory nature, and high compressive strength. Though less common in mainstream stabilization, it has been shown to significantly improve strength, especially the California Bearing Ratio [4]. Its effectiveness may vary based on local soil and brick properties [5].

Clay brick dust (CBD), a by-product of brick kiln operations, contains high levels of silica (55%), aluminum oxide (15%), iron oxide (8%), calcium oxide (7%), magnesium oxide (2%), and sulphur trioxide (1%) [6]. These oxides promote pozzolanic reactions, enhance particle bonding, and improve strength [7]. While CBD contains crystalline silica, which can pose respiratory risks, it is considered safe under controlled construction conditions. Importantly, CBD reuse supports sustainable construction by reducing construction and demolition waste, preserving resources, and aligning with circular economy goals [8].

In addition to brick dust, other construction-derived powders such as marble dust, stone dust, and granite dust have also been investigated for soil stabilization. For instance, [9] reported that waste marble dust reduced swelling and plasticity while significantly improving California Bearing Ratio (CBR) and Unconfined Compressive Strength (UCS) of expansive clays. A recent study by [10] confirmed similar enhancements in geotechnical properties with marble dust inclusion. Similarly, [11] demonstrated that the addition of stone dust (10–50%) improved the maximum dry density (MDD) and CBR of clayey soils, while reducing the optimum moisture content (OMC). Broader evaluations, such as those by [12], examined multiple construction dusts (stone, granite, marble, and tile powders) and reported significant benefits in compaction, shear strength, and durability. These findings suggest that a variety of locally available construction by-products can serve as sustainable stabilizers, aligning with circular economy practices and environmentally friendly soil improvement strategies.

In prior studies, clay soil with 5% to 30% CBD showed improved dry density and reduced plasticity. For example, dry density increased from 1.72 g/cc to 1.89 g/cc with 5% CBD, with optimum water content at 19.52%. A 10% CBD mix yielded a density of 1.87 g/cc at 17.64% moisture [13]. Swathi (et al.) reported reduced swelling and shrinkage in expansive soils at 50% CBD, while [5] observed the highest dry density and CBR at 15% CBD, especially when combined with lime. These suggest that lower CBD ratios can yield meaningful improvements.

CBD has also been evaluated in blended stabilizers. Khan (et al.) showed that CBD combined with cement and fly ash in a 1:3 ratio enhanced unconfined compressive strength and reduced shrinkage, meeting AASHTO specifications. Gupta (et al.) tested CBD with rubber fiber and calcium chloride, showing strength gains depending on curing time, cement content,

https://doi.org/10.31436/iiumej.v26i3.3821

and compaction. Rubber fibers improved shear strength but had little impact on liquid limit, whereas calcium chloride in CBD reduced the liquid limit significantly.

Although prior studies support CBD as a sustainable stabilizer, results vary based on soil type, dosage, and additive combinations. High percentages offer performance gains but may not be cost-effective. Although prior studies support CBD as a sustainable stabilizer, most existing research has focused on blended stabilizers such as brick dust with cement, lime, or other additives to enhance pozzolanic reactivity and durability [16]. However, limited studies have isolated the effect of CBD as a sole stabilizing agent, particularly in clayey sand applications. Given the environmental benefits and widespread availability of brick waste, further investigation into its standalone performance is essential to validate its practical and sustainable application.

This study investigates the impact of 15%, 20%, and 25% CBD on the physical and mechanical properties of clayey sand. Laboratory tests include particle size analysis, Atterberg limits, moisture content, compaction, and California Bearing Ratio to evaluate CBD's potential as a standalone stabilizer. The findings aim to support sustainable soil improvement using construction waste in geotechnical engineering.

2. METHODOLOGY

2.1. Sample Collection

Clayey sand soil samples (Fig. 1a) were collected from a site near the riverside of Sungai Gombak, Gombak, Malaysia (3°15'21"N, 101°44'30"E), a location known for problematic soil profiles (fig. 2). Brick dust, used as the stabilizing agent, was sourced from the IIUM construction yard. The CBD (Fig. 1b) was air-dried, manually crushed, and sieved to ensure a consistent particle size and avoid oversized aggregates that could affect the mixing and compaction process.

The samples were prepared by mixing the clayey sand with 0% (control), 15, 20, and 25% CBD by weight. This allowed a comparative evaluation of different proportions to identify the most effective dosage.

(a) Clayey Sand (b) CBD

Figure 1. Samples of (a) Clayey sand and (b) CBD

https://doi.org/10.31436/iiumej.v26i3.3821

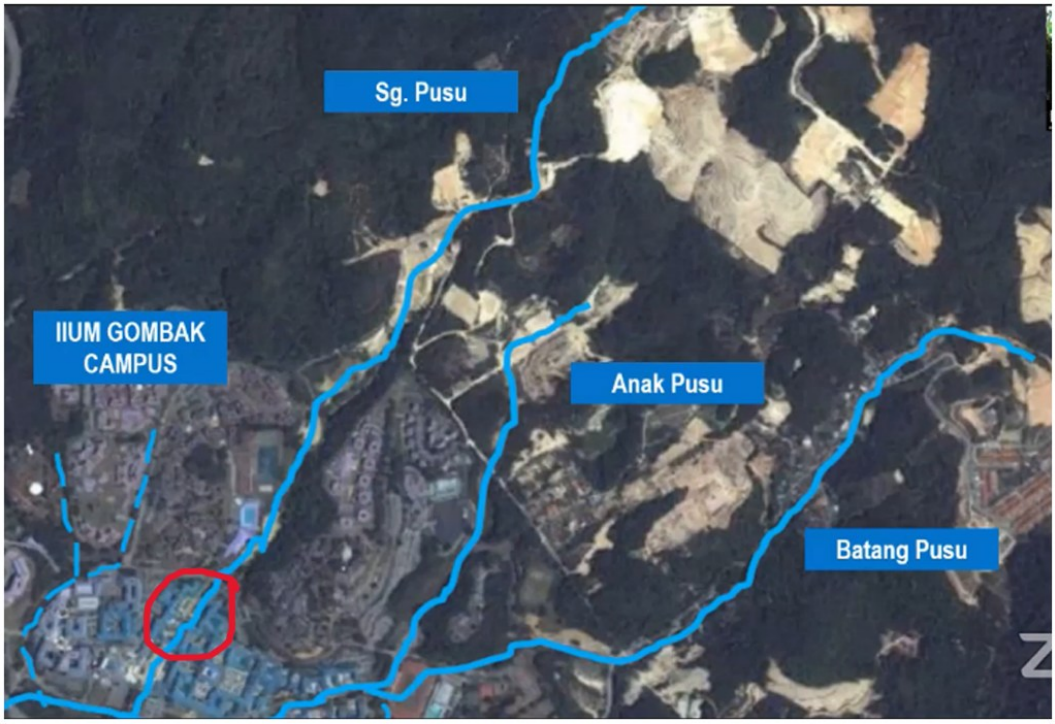


Figure 2. Location of soil sample collected at Sungai Gombak, Malaysia [19]

2.2. Atterberg limit test

The Atterberg limits test was conducted to determine the liquid limit (LL), plastic limit (PL), and plasticity index (PI), which are essential for evaluating soil consistency and its behaviour under varying moisture conditions. All procedures followed the ASTM D4318 standard.

Four soil mixtures were prepared: untreated soil (0% CBD) and soils stabilized with 15, 20, and 25% clay brick dust (CBD) by weight. The samples were air-dried and sieved through a No. 40 sieve (425 µm) to ensure uniform particle size.

For the plastic limit (PL) determination, each moist soil sample was hand-rolled on a smooth glass plate into threads approximately 3 mm in diameter. Rolling continued until the threads crumbled at around 3.2 mm, indicating the transition from the plastic to semi-solid state. The crumbled samples were placed in moisture-proof containers, weighed, and then oven-dried at 105 ± 5°C for 24 hours. The moisture content at the point of disintegration was recorded as the PL.

For the liquid limit (LL), the Casagrande cup method was employed. A soil paste was placed in the brass cup, and a groove was formed at the centre using a standard grooving tool. The cup was dropped from a height of 10 mm at a rate of two drops per second. The number of blows required to close the groove over a distance of approximately 13 mm was recorded. The LL was determined from the flow curve corresponding to 25 blows.

The plasticity index (PI) was calculated as the difference between the LL and PL (PI = LL − PL). The PI reflects the range of moisture content within which the soil remains plastic and workable. A reduction in PI after CBD treatment generally indicates improved soil behaviour, particularly through reduced shrink-swell potential and enhanced dimensional stability.

2.3. Standard Proctor Test

The compaction characteristics of both untreated and stabilized soil samples were evaluated using the Standard Proctor Test in accordance with ASTM D698, to determine their optimum moisture content (OMC) and maximum dry density (MDD). For each test, approximately 5 kilograms of air-dried soil were thoroughly mixed with clay brick dust (CBD) at proportions of 0%, 15%, 20%, and 25% by weight. The required amount of distilled water was added gradually to each mixture to achieve the desired moisture content. Each mix was then compacted in a standard cylindrical mould in three equal layers. Each layer received 25 blows from a standard Proctor rammer dropped from a height of 12 inches (305 mm), ensuring uniform compaction throughout the sample. After compaction, the collar was removed, and the soil surface was leveled with a steel straightedge.

The mass of the mould and compacted soil was measured to determine the bulk (wet) density. Small samples were then extracted from the top and bottom of each mould and placed in moisture containers. These containers were weighed and then oven-dried at 105 ± 5°C for 24 hours to determine their moisture content. Using the wet density and moisture content data, the dry density of each specimen was calculated. This process was repeated with increasing moisture content (typically in 2–3% increments) until a peak in dry density was observed. The point at which the dry density reached its maximum was recorded as the MDD, and the corresponding moisture content was taken as the OMC. These values were used to assess how the inclusion of CBD influenced the compaction behavior of clayey sand, which is critical for These values were used to assess how the inclusion of CBD influenced the compaction behaviour of clayey sand, which is critical for field performance in roadwork and foundation applications. In roadwork, achieving an optimum moisture content and high dry density is essential to ensure subgrade stability and to prevent rutting or settlement under traffic loads. For foundation applications, well-compacted soil provides uniform bearing support, reduces differential settlement, and enhances load distribution to prevent structural failure. Therefore, understanding the compaction characteristics is vital for determining the suitability of CBD-treated soils in these geotechnical scenarios.

2.4. California Bearing Ratio Test (CBR)

The bearing capacity of the soil mixtures was evaluated using the California Bearing Ratio (CBR) test, as specified in ASTM D1883. This test measures the resistance of a compacted soil sample to penetration by a standard piston under controlled loading conditions, simulating the performance of subgrade in flexible pavement design.

The soil-CBD mixtures at 0, 15, 20, and 25% by weight were compacted into CBR moulds, which differ in size from the Standard Proctor mould. The CBR moulds used in this study had an internal diameter of 152 mm and a height of 178 mm. Although the mould dimensions varied, the compaction procedure adhered to the same method, whereby the samples were compacted in three equal layers with 25 blows per layer using the standard Proctor rammer.

Following compaction, the specimens were cured for 24 hours and then fully submerged in water for a 4-day soaking period. This duration was selected to simulate prolonged saturation conditions that are typically encountered in weak subgrade environments, particularly during periods of heavy rainfall or poor drainage. The soaking time was based on previously published studies, including [20], where a four-day soaking period was used to evaluate the saturated strength behaviour of stabilized soils.

After soaking, the samples were mounted on the CBR testing machine. A metal plunger with a cross-sectional area of 1,950 mm² was then driven into the centre of each sample at a

https://doi.org/10.31436/iiumej.v26i3.3821

constant rate of 1.25 mm/min. During the penetration process, load values were recorded at incremental depths ranging from 0.0 mm to 7.5 mm. The CBR value at each depth was calculated by comparing the measured load to standard reference loads (13.24 kN at 2.5 mm and 19.96 kN at 5.0 mm penetration). The higher of the two calculated CBR values was taken as the final CBR for each sample. Post-test, the moisture content of the tested samples was also determined by oven drying. The test results were used to compare the load-bearing improvement of soil treated with different percentages of CBD, where higher CBR values indicate enhanced strength and suitability for pavement subgrade construction.

3. RESULTS AND DISCUSSION

3.1. Particle Size Distribution

The particle size distribution test was conducted to classify the soil type and evaluate its grading characteristics. To ensure the reproducibility of results, the test was conducted on three representative soil samples. The percentage passing values obtained from each sieve size were then averaged to plot the representative particle size distribution curve, in accordance with ASTM D422 and BS 1377. This approach minimizes the effect of variability between samples and provides a reliable gradation profile for subsequent analysis. The collected soil was subjected to two laboratory tests: sieve analysis and hydrometer analysis. The results of both tests were combined to develop the particle size distribution curve, which illustrates the relative proportions of gravel, sand, silt, and clay. The resulting curve for the parent soil is shown in Fig.3.

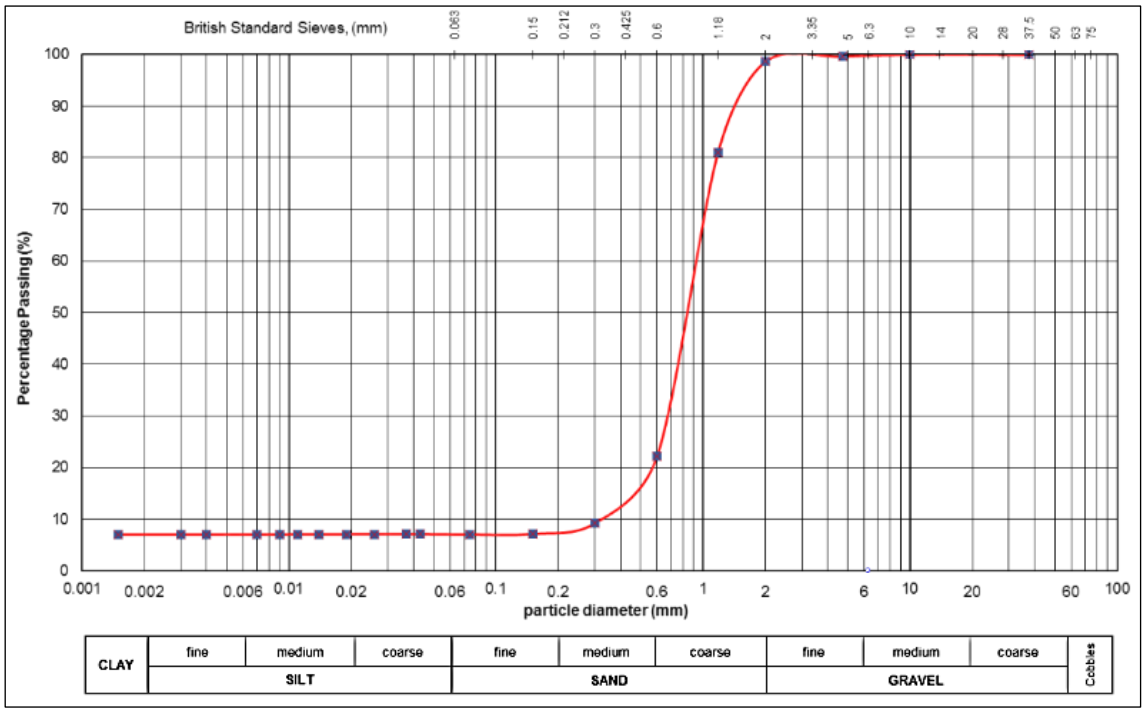


Figure 3. Particle size distribution graph

As shown in Table 1, the soil consisted predominantly of sand, which made up 84.54% of the sample. This was followed by clay (6.99%), gravel (5.51%), and silt (2.96%). According to the Unified Soil Classification System (USCS), the soil can be classified as clayey sand (SC) since more than 50% of the sample was retained on the No. 200 sieve, which has an opening size of 75 microns, and it contained a significant proportion of fines. According to the USCS,

the clay fraction is defined as particles ≤0.002 mm (hydrometer), silt as 0.002–0.075 mm, sand as 0.075–4.75 mm, and gravel as >4.75 mm. The #200 sieve (75 μm) separates fines (silt + clay) from sand, but the clay fraction is determined specifically by hydrometer analysis. The similar proportions of clay (6.99%) and gravel (5.51%) reflect the residual nature of the soil, where coarse fragments coexist with fine particles; however, the dominance of sand (84.54%) still governs the overall classification as a clayey sand (SC).

Table 1. Particle size distribution

Type of particle	Percentage from total amount [%]
Gravel (4.75 mm – 76.2 mm)	5.51
Sand (0.075 mm – 4.75 mm)	84.54
Silt (0.002 mm – 0.075 mm)	2.96
Clay (≤ 0.002 mm)	6.99

The sieve analysis was conducted using a standard stack of sieves, including the No. 4 (4.75 mm), No. 10 (2 mm), No. 40 (425 μm), and No. 200 (75 μm) sieves, to ensure accurate classification of gravel, sand, silt, and clay fractions.

The particle size curve further supports the conclusion that the soil is poorly graded, as indicated by the calculated coefficients of uniformity (Cu = 5.81) and gradation (Cc = 0.86). These values suggest limited gradation and reduce interlocking potential among particles. As a result, the soil may exhibit poor compaction behaviour and increased susceptibility to water infiltration and settlement. The introduction of finer materials such as brick dust is therefore expected to fill the voids between sand particles, improving particle packing and enhancing the engineering properties of the soil.

3.2. Atterberg Limits Test

The Atterberg limits test was performed to assess the consistency limits and plasticity characteristics of the soil before and after stabilization with clay brick dust (CBD). The results, presented in Table 2, show a significant reduction in both the liquid limit (LL) and plastic limit (PL) as the percentage of CBD increased. The untreated soil sample (CS-CBD0) exhibited an LL of 63.9%, a PL of 41.4%, and a plasticity index (PI) of 22.5, classifying it as a high plasticity soil.

When 15% CBD was introduced, the LL and PL decreased to 44.2% and 27.7%, respectively, resulting in a PI of 18.6, which still falls within the intermediate plasticity range. At 20% CBD, further reductions to 44.0% (LL) and 26.5% (PL) produced a PI of 17.5, maintaining the same intermediate classification. However, at 25% CBD, the LL and PL decreased significantly to 25.5% and 9.8%, respectively, resulting in a PI of 15.7, which can be classified as low plasticity. These findings are consistent with prior research using various construction by-products. [21] and [22] reported reductions in plasticity index in expansive soils treated with marble dust. SAS Yaba also found that adding 5–10% ceramic dust to clay soils significantly decreased LL, PL, and PI. At the same time, [24] observed a steady reduction in PI, which almost halved at a 30% ceramic tile dust content. Anil (et al.) reported improvements in compaction, CBR, and index properties, including PI, when tile waste was used as a stabilizer. Similarly, [18] highlighted enhanced volumetric stability and lowered plasticity using marble dust. The behaviour of CBD in our study, which yielded a PI of 15.7 at a 25% dosage and classified the soil as low plasticity, is therefore supported by these parallel findings with other stabilizing powders.

This consistent downward trend in PI values across CBD-treated samples indicates that the inclusion of CBD effectively reduces the soil's plastic behaviour. The replacement of clay particles with non-plastic fines from CBD weakens the cohesive bonds within the soil matrix and enhances its workability. Lower plasticity is generally associated with better dimensional stability and reduced shrink–swell potential, both of which are desirable for construction and pavement applications.

Table 2. Atterberg limit results

Soil sample	Liquid limit	Plastic limit	Plasticity Index
CS-CBD0	63.9	41.4	22.5
CS-CBD15	44.2	27.7	18.6
CS-CBD20	44.0	26.5	17.5
CS- CBD25	25.5	9.8	15.7

3.3. Standard Proctor Compaction Test

The compaction characteristics of the soil samples were evaluated using the Standard Proctor Test. The optimum moisture content (OMC) and maximum dry density (MDD) for each sample were evaluated. The control sample (Soil + 0% CBD) exhibited an OMC of 5.8% and an MDD of 2.30 g/cm³. The addition of 15% brick dust increased the OMC to 6.4% and the MDD to 2.40 g/cm³, suggesting enhanced compaction due to better particle rearrangement and reduced water absorption.

However, at 20% CBD content, the OMC further increased to 6.5%, while the MDD dropped to 2.23 g/cm³. This reduction in density may be attributed to the presence of excessive fines, which can hinder effective compaction. Interestingly, the sample with 25% brick dust showed a lower OMC of 5.31% and a slight recovery in MDD to 2.35 g/cm³. These fluctuations highlight the importance of optimizing the quantity of stabilizers, as excessive brick dust can disrupt soil structure and compaction behavior.

Fig. 4 illustrates the variation in OMC with increasing CBD content. The OMC increased with the addition of CBD up to 20% but dropped to 25%. This indicates that, up to a certain level, the soil requires more water for compaction due to its high absorption capacity.

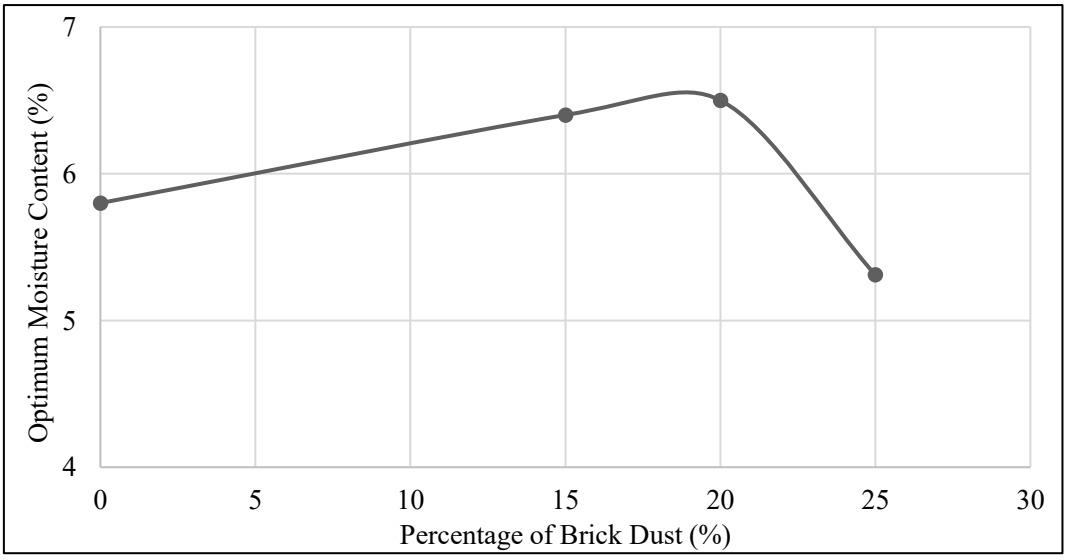


Figure 4. Optimum moisture content graph

https://doi.org/10.31436/iiumej.v26i3.3821

Fig. 5 presents the trend in MDD values as the CBD content increases. Initially, MDD increases by 15% with 15% CBD due to improved packing and interlocking between particles. However, MDD decreases at 20% CBD, likely due to the excessive fine material disrupting optimal compaction. At 25% CBD, MDD increases again, possibly due to re-compaction of finer particles and better particle orientation during ramming. It is possible that increasing CBD content beyond 25%, such as to 30%, may lead to a further decline in MDD due to the excessive presence of non-cohesive fine particles. These fines may prevent proper particle interlocking and reduce the soil's overall compactability. Over-stabilization with brick dust can create a loosely packed matrix with higher void ratios, thus reducing dry density. However, this hypothesis requires experimental validation and may be explored in future work.

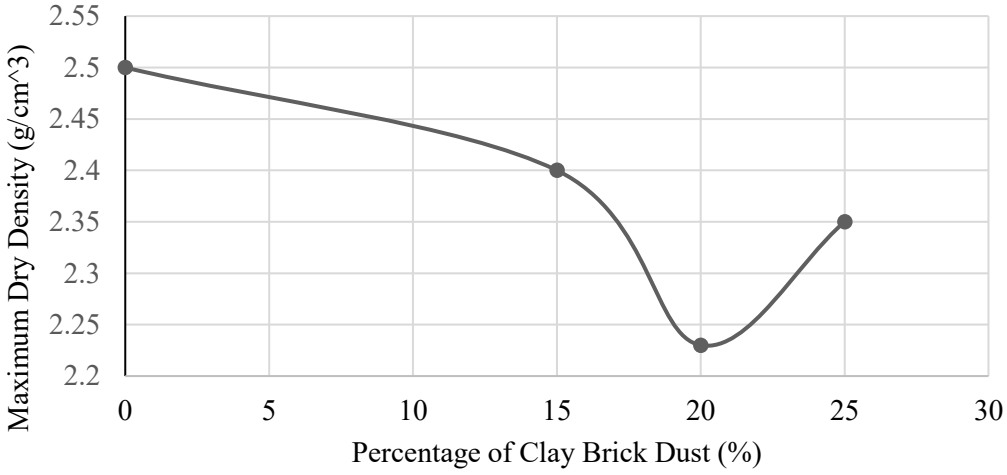


Figure 5. Maximum dry density graph

To better understand the overall behavior of soil with CBD under compaction, Fig. 6 shows the dry density versus moisture content curve for all mixtures. The curve demonstrates the typical compaction behavior where dry density increases with moisture content until a peak point is reached and then begins to decline. For the untreated soil, the peak dry density occurs at a lower moisture content (around 4.5%), while for CBD-treated samples, the peak shifts slightly, with lower MDDs observed at higher moisture contents. This shift supports the observation that CBD affects the water-holding and densification behavior of the soil.

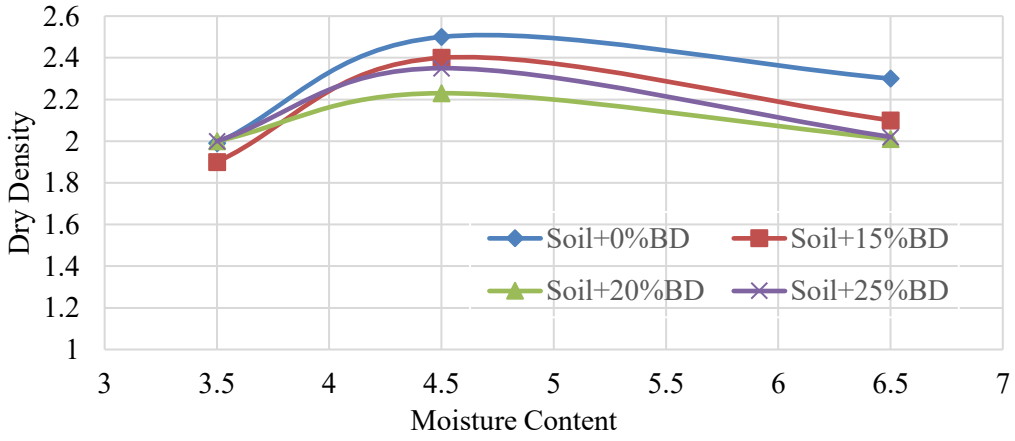


Figure 6. Dry Density vs Moisture Content Graph

https://doi.org/10.31436/iiumej.v26i3.3821

In conclusion, the data from the Standard Proctor test indicate that while brick dust improves compaction characteristics up to a certain dosage, an excessive amount can reduce efficiency. The 15% CBD mixture yielded the highest MDD, but the 20% mixture offered the best balance between OMC and MDD, making it the most suitable dosage overall.

3.4. California Bearing Ratio (CBR) Test

Fig. 7 illustrates the relationship between load applied to the plunger and the corresponding penetration depth during the CBR tests for soil samples containing 0, 15, 20, and 25% CBD. This load-penetration curve provides critical insight into the mechanical response of the soil under simulated loading conditions.

From the graph, it is evident that as the CBD content increases, the soil's resistance to penetration also improves. This indicates that while brick dust improves soil strength, excessive amounts may start to negatively affect load transfer due to overfilling of voids and loss of particle interlock.

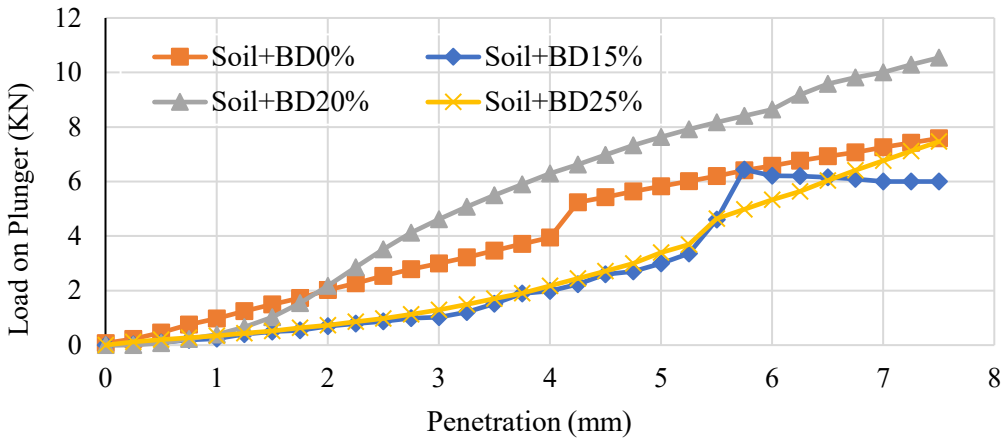


Figure 7. Load on plunger vs Penetration graph

The California Bearing Ratio (CBR) test was carried out to assess the load-bearing capacity of the clayey sand soil treated with different proportions of clay brick dust (CBD). The CBR values obtained from the test are shown in Table 4. The control sample (Soil + 0% CBD) recorded a CBR value of 27.0%, which reflects the natural strength of the untreated clayey sand. Upon introducing 15% CBD into the soil, the CBR value increased slightly to 28.6%, suggesting a modest improvement in strength. This increase may be attributed to improved grain interlocking and reduced moisture sensitivity resulting from the presence of finer, angular CBD particles.

The highest improvement in CBR was observed with the 20% CBD mixture, which achieved a CBR value of 30.0%. This indicates that 20% brick dust is the optimum content for enhancing soil strength in this study. The improvement is likely due to the optimal balance between filler content and compaction efficiency, resulting in improved stress distribution underload. However, when the CBD content was increased to 25%, the CBR value slightly declined to 29.5%, which may be due to the excess fines interfering with load transfer between soil particles. These results underscore the importance of determining the optimal dosage of stabilizer to achieve the best mechanical performance.

Although the 20% CBD sample exhibited a reduction in maximum dry density (MDD) and an increase in optimum moisture content (OMC), it achieved the highest CBR value at 30%.

https://doi.org/10.31436/iiumej.v26i3.3821

This seemingly contradictory behaviour can be explained by the enhanced pozzolanic reactivity and filler effect of brick dust. The increased fine content at this dosage promotes better particle packing and a denser matrix under penetration load, even if the dry density is lower. Additionally, the presence of reactive oxides such as silica, alumina, and calcium oxide in CBD facilitates cementitious bonding in moist conditions. The higher moisture content in the 20% mix likely enabled improved hydration and pozzolanic reactions, which contributed to better bonding between soil particles. This enhanced interparticle bonding leads to higher resistance to deformation under load, thus increasing the CBR value. Therefore, while compaction density decreased slightly, the overall load-bearing performance improved due to microstructural enhancements provided by CBD at 20%.

The CBR values for each soil-brick dust mixture were evaluated. The trend shows a general increase in strength with increasing CBD content up to 20%, followed by a slight decrease to 25%. This pattern supports the conclusion that over-stabilization can be counterproductive beyond a certain point.

In addition to the tabulated data, Fig. 8 illustrates the trend of CBR values across the different samples. From the graph, it is visually evident that the sample with 20% CBD had the highest CBR value. The slope of the curve increases from 0% to 20% CBD and slightly declines at 25%, confirming that the soil's strength increases up to an optimum level and then begins to plateau or slightly decreases. For instance, [10] reported that marble dust stabilization improved soil properties, with the highest CBR values achieved at the optimum dosage. Similarly, [11] observed that stone dust reinforcement produced significant gains in CBR.

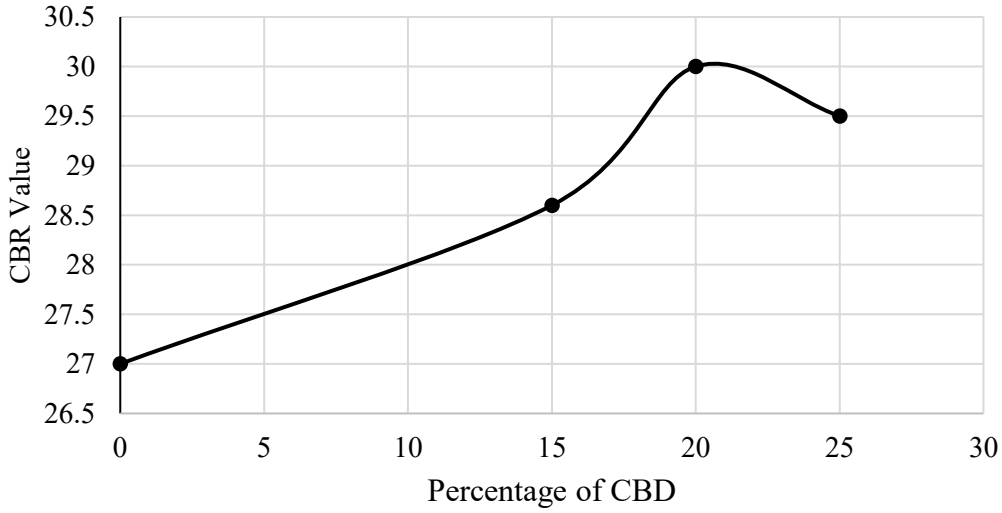


Figure 8. California Bearing Ratio values

Despite the observed increase in optimum moisture content (OMC) and decrease in maximum dry density (MDD) at 20% CBD addition, the CBR value improved significantly. This apparent contradiction is due to the pozzolanic and filler effects of CBD, which enhance bonding between particles and increase stiffness under load. The finer particles in CBD fill the voids between sand grains, resulting in a denser particle matrix that improves resistance to penetration, even if the compacted dry density is slightly lower. Additionally, higher moisture content may facilitate hydration and pozzolanic reactions, further improving the soil's load-bearing capacity. Thus, 20% CBD provides optimal subgrade strength despite not achieving the highest dry density.

4. CONCLUSION

This study investigated the effects of incorporating varying proportions of CBD on the physical and mechanical properties of clayey sand soil. Based on the results obtained from laboratory testing, including particle size distribution, Atterberg limits, Standard Proctor compaction, and California Bearing Ratio (CBR) tests. It can be concluded that brick dust has significant potential as a sustainable soil stabilizing agent.

The addition of CBD consistently reduced the plasticity of the soil, with the plasticity index (PI) decreasing from 22.5 (untreated) to 15.7 at a 25% CBD concentration. This reduction in plasticity indicates improved dimensional stability and lower susceptibility to volume change. The compaction results showed that the 15% CBD mixture achieved the highest maximum dry density (2.40 g/cm³). In comparison, the 20% CBD mixture offered a balanced outcome in terms of optimum moisture content and density. The CBR results further confirmed this, with 20% CBD yielding the highest bearing capacity of 30.0%, compared to 27.0% for untreated soil.

Overall, the findings suggest that the inclusion of 20% brick dust by weight provides the optimal improvement in strength performance, as evidenced by the highest CBR value. Although the 15% CBD mixture showed slightly higher dry density, the 20% CBD mixture offered a more favourable balance between compaction efficiency and mechanical strength. This is attributed to improved pozzolanic reactivity, better interparticle bonding, and the filler effect of the finer CBD particles, which enhance load-bearing capacity despite a marginal decrease in dry density.

The absence of chemical and mineralogical characterisation of both the clayey sand and brick dust is acknowledged as a limitation of this study. Future research will incorporate microstructural and chemical analyses, such as SEM, XRD, and XRF, to better understand the interaction mechanisms and validate the observed improvements in strength and plasticity.

REFERENCES

[1] N. A. Saparudin, N. Kasim, K. A. Taib, W. N. A. W. Azahar, N. A. Kasim, and M. Ali, "Improvement of Problematic Soil Using Crumb Rubber Tyre," *IIUM Engineering Journal*, vol. 23, no. 2, pp. 72–84, 2022, doi: 10.31436/IIUMEJ.V23I2.2293.

[2] R. Ventini, M. Pirone, and C. Mancuso, "Soil water retention curves of a silty clayey sand compacted at different dry density," *E3S Web Conf.*, vol. 382, 2023, doi: 10.1051/e3sconf/202338209007.

[3] T. S. Krishna, N. S. Basha, and K. S. Chamberlin, "Experimental study on black cotton soil to be used as filling material by strenghtning with baggasse ash and brick dust," *Int. J. Recent Technol. Eng.*, vol. 7, no. 6C2, pp. 700–704, 2019.

[4] Y. Salimah, A., Hazmi, M., Fathur Rouf Hasan, M., Agung, P. A. M., "Soil stabilization using redbrick powder toward California Bearing Ratio.," 2022.

[5] A. Salimah, M. Hazmi, M. Fathur Rouf Hasan, P. A. M. Agung, and . Yelvi, "A comparative study of red brick powder and lime as soft soil stabilizer," *F1000Research*, vol. 10, p. 777, 2021, doi: 10.12688/f1000research.27835.1.

[6] S. Paudel and B. Kumar Dahal, "Effect of Brick Dust on Soil and Strength Improvement with the use of Plastic Waste," vol. 8914, pp. 197–201, 2022.

[7] R. A. Blayi, A. F. H. Sherwani, H. H. Ibrahim, and S. J. ACBDullah, "Stabilization of high-plasticity silt using waste brick powder," *SN Appl. Sci.*, vol. 2, no. 12, 2020, doi: 10.1007/s42452-020-03814-8.

[8] A. Al-Baidhani, A. Al-Taie, and M. S. Student, "Review of Brick Waste in Expansive Soil Stabilization and Other Civil Engineering Applications," *J. Geotech. Stud.*, vol. 4, no. 3, pp. 2581–9763, 2019.

[9] Saygili, A. (2015). *Use of Waste Marble Dust for Stabilization of Clayey Soil.* International Journal of Civil Engineering and Technology, 6(3), 22–30.

[10] Ali, R., Khan, M., & Farooq, A. (2025). *Sustainable soil stabilization using marble dust waste on high-plasticity soils: Physical and mechanical properties study.* Construction and Building Materials, 412, 133123. https://doi.org/10.1016/j.conbuildmat.2025.133123

[11] Khan, A.S.G., & Prabu, N. (2020). *Effect of Stone Dust on Geotechnical Properties of Soil.* International Journal of Technical Innovation in Modern Engineering & Science, 6(2), 45–51.

[12] Gupta, R., Sharma, S., & Das, B. (2024). Stabilization of soils using construction dusts (stone, granite, marble, and tiles): A comparative study. Journal of Geotechnical Studies, 4(3), 2581–9763.

[13] A. S. G. Khan, N. A., & Prabu, "Structural design of fuselage," *Int. J. Tech. Innov. Mod. Eng. Sci.*, 2020.

[14] A. P. K. G. Prema Swathi, Ch. Srinivas, Kodali Nagaraju, Pentakota Shyam Sunda, Pentakota Shyam Sundar, Dunna Duryodhana, "Stabilization of Expansion Soil Using Brick Dust," *INTERANTIONAL J. Sci. Res. Eng. Manag.*, 2025.

[15] S. K. Mahto, S. Sinha, and A. Kumar, "Mechanical Properties, Shrinkage and Microstructural Evaluation of Stabilised Intermediate Silty Clay for Low-Volume Roads," *Arab. J. Sci. Eng.*, vol. 49, no. 4, pp. 4675–4684, 2024, doi: 10.1007/s13369-023-08169-x.

[16] M. Tech Scholar, A. Professor, and I. , "Enhancement in Properties of Clayey Soil with addition of Calcium Chloride, Brick Kiln Dust and Rubber Fiber," no. July, pp. 213–221, 2023, [Online].

[17] H. Jangde and F. Khan, "Comment on Energy-Efficient Alternative for Different Types of Traditional Soil Binders," Stud. Geotech. Mech., vol. 45, no. 1, pp. 72–87, 2023, doi: 10.2478/sgem-2022-0029.

[18] V. Narnoli, S. Suman, and R. Kumar, "Stabilization of Alluvial Soil Using Marble Dust, Lime and Burnt Brick Dust for Road Construction," Indian Conf. Geotech. GEO-ENVIRONMENTAL Eng., vol. 5, no. May, pp. 1–6, 2019, [Online]. Available: https://www.researchgate.net/publication/333488726_STABILIZATION_OF_ALLUVIAL_SOIL_USING_MARBLE_DUST_LIME_AND_BURNT_BRICK_DUST_FOR_ROAD_CONSTRUCTION

[19] F. A. Razali and M. F. Ayob, "Appraisal of Cost-Effective Approaches for the Pusu River Rehabilitation," *J. Archit. Plan. Constr. Manag.*, vol. 12, no. 2, pp. 40–57, 2022, doi: 10.31436/japcm.v12i2.669.

[20] D. Melese, B. Aymelo, T. Weldesenbet, and A. Sorsa, "Utilization of Black Cotton Soil Stabilized With Brick Dust-Lime for Pavement Road Construction: an Experimental and Numerical Approach," *Balt. J. Road Bridg. Eng.*, vol. 18, no. 4, pp. 42–64, 2023, doi: 10.7250/bjrbe.2023-18.618.

[21] Scabat, R. & Nanda, R. (2011). Stabilization of expansive soil using marble dust. *International Journal of Civil Engineering and Technology*, 2(3), 45–52.

[22] Baser, T. (2009). Effect of marble dust on engineering properties of clayey soil. *Construction and Building Materials*, 23(5), 2187–2191.

[23] Yaba, S.A.S. (2021). Effect of ceramic dust on the index properties of clay soils. *Nigerian Journal of Technology*, 40(2), 289–297.

[24] Chong, S.Y., Lim, K.S., & Rahman, M.A. (2022). Stabilization of tropical residual soil using ceramic tile dust. *E3S Web of Conferences*, 325, 03005.

[25] Anil, U., Kumar, P., & Rao, S. (2020). Utilization of tile waste for soil stabilization. *International Journal of Engineering Research and Technology*, 9(4), 554–559.

https://doi.org/10.31436/iiumej.v26i3.3825

NUMERICAL INVESTIGATION OF SEDIMENT SCOUR DOWNSTREAM OF ADJUSTABLE AND FIXED SUBMERGED WEIR USING FLOW-3D

SITI NURAIN CHE MOHD AZMI, SAERAHANY LEGORI IBRAHIM*, IZIHAN IBRAHIM, DANI IRWAN MASBAH

Civil Engineering Department, Kulliyah of Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia

**Corresponding author: saerahany@iium.edu.my*

(Received: 2 July 20205; Accepted: 27 July 2025; Published online: 9 September 2025)

ABSTRACT: Scouring downstream of hydraulic structures poses significant risks to structural stability and sediment transport processes in river systems. This study presents a numerical investigation of sediment scour behaviour downstream of two submerged weir types, adjustable submerged weirs (hydraulic tilting gates) and fixed submerged weirs (Crump weirs), using FLOW-3D software. Field measurements from Acoustic Doppler Current Profiler (ADCP) surveys were used to validate the hydrodynamic model, with a finer mesh size of 1.5 m achieving high accuracy (R^2 = 0.998, RMSE = 0.018 m). This mesh size was subsequently adopted for sediment transport simulations. The research focuses on evaluating velocity at the bed and scour depth under a range of flow discharge series, from 183.30 m^3/s to 3666.00 m^3/s, over a 2-year ARI for the Kelantan River, across 20 simulation runs, with low, moderate, and high discharges, classified using a percentile-based approach. The simulations adopt the Reynolds-Averaged Navier-Stokes (RANS) equations with the RNG turbulence model, coupled with the Meyer-Peter and Müller sediment transport formulation. Results reveal that the adjustable submerged weir generally produces higher bed-level velocities at low to moderate discharges, indicating a tendency toward more concentrated local scour. This is likely due to the flow being constricted and accelerated as it passes over and through the adjustable structure, thereby increasing shear stress on the bed. In contrast, the fixed Crump weir exhibits significantly higher peak velocities under extreme flow conditions, leading to broader but less focused erosion patterns, likely because its fixed geometry causes the flow to spread out more evenly downstream, reducing localised acceleration but increasing the overall area affected by erosion. Scour depth analysis reveals that the adjustable weir induces deeper scour holes, with a maximum depth of -5.917 m, whereas the fixed weir reaches a maximum depth of -4.128 m. These outcomes highlight the influence of structural geometry and operational flexibility on downstream hydraulic behaviour and sediment response.

ABSTRAK: Hakisan di hilir struktur hidraulik boleh menjejaskan kestabilan struktur dan proses pengurusan sedimen sungai. Kajian ini menjalankan analisis berangka terhadap keadaan hakisan sedimen pada dua jenis empangan tenggelam iaitu empangan tenggelam boleh laras (pintu hidraulik condong) dan empangan tenggelam kekal (empangan Crump), menggunakan perisian FLOW-3D. Pengukuran lapangan menggunakan Doppler Akustik Pembukah Semasa (ADCP) digunakan bagi mengesahkan model hidrodinamik, di mana saiz mesh lebih halus 1.5 m mencapai ketepatan tinggi (R^2 = 0.998, RMSE = 0.018 m). Seterusnya digunakan bagi simulasi pengangkutan sedimen. Kajian ini memfokuskan pada penilaian halaju pada dasar sungai dan kedalaman hakisan bagi beberapa siri pelepasan aliran sungai berkelajuan antara 183.30 m^3/s hingga 3666.00 m^3/s selama tempoh 2 tahun (ARI) di Sungai Kelantan. Ia melibatkan 20 siri simulasi, dengan aliran rendah, sederhana dan tinggi yang

https://doi.org/10.31436/iiumej.v26i3.3825

diklasifikasikan menggunakan pendekatan berasaskan persentil. Simulasi ini menggunakan persamaan Navier-Stokes Purata-Reynolds (RANS) dengan model pergolakan RNG, digabungkan bersama rumusan Meyer-Peter dan Müller bagi pengangkutan sedimen. Dapatan kajian menunjukkan empangan tenggelam boleh laras menghasilkan halaju dasar lebih tinggi pada aliran rendah hingga sederhana, mendorong kepada hakisan setempat yang bertumpu. Hal ini berkemungkinan disebabkan oleh aliran yang tersekat dan dipercepatkan ketika melalui struktur boleh laras, lalu meningkatkan tegasan ricih pada dasar sungai. Sebaliknya, empangan Crump jenis kekal menunjukkan halaju puncak lebih tinggi semasa aliran deras, menghasilkan corak hakisan yang lebih meluas tetapi kurang bertumpu, kerana geometri kekal menyebabkan aliran tersebar lebih sekata di hilir, sekaligus mengurangkan pecutan setempat, tetapi memperluas kawasan yang terjejas oleh hakisan. Analisis kedalaman hakisan mendedahkan bahawa empangan boleh laras menghasilkan lubang hakisan yang lebih dalam, dengan kedalaman maksimum -5.917 m, manakala empangan kekal mencapai kedalaman maksimum -4.128 m. Dapatan ini menekankan pengaruh geometri struktur dan fleksibiliti operasi pada keadaan hidraulik hilir serta respon sedimen.

KEY WORDS: *Scouring, Submerged Weir, Hydraulic Tilting Gate, Crump Weir, FLOW-3D*

1. INTRODUCTION

1.1. Scouring

Scouring refers to the erosion process caused by flowing water that removes sediment from the riverbed [1], often leading to significant changes in bed morphology downstream of hydraulic structures such as submerged weirs. This phenomenon is critical because excessive scour can compromise the stability and safety of these structures, potentially resulting in structural failure and associated economic losses [2]. Numerous studies have investigated scour downstream of submerged weirs, focusing on various factors influencing scour depth and geometry, including flow conditions, sediment characteristics, and weir design.

Previous research has demonstrated that the shape and height of weirs significantly affect the development of scour holes. For example, experimental studies comparing different weir geometries have shown that rectangular weirs tend to produce 36-56% deeper scour holes compared to trapezoidal weirs under similar flow conditions, with near-bed velocities of 2-3 m/s that generate 15 N/m², intensifying the scouring process [3]. Additionally, the flow regime, whether free-flow or submerged, also plays a critical role. Based on Figure 1, submerged weirs under live bed conditions exhibit fluctuating scour depths due to migrating bedforms, with equilibrium depths 10% lower than clear water scour scenarios [4].

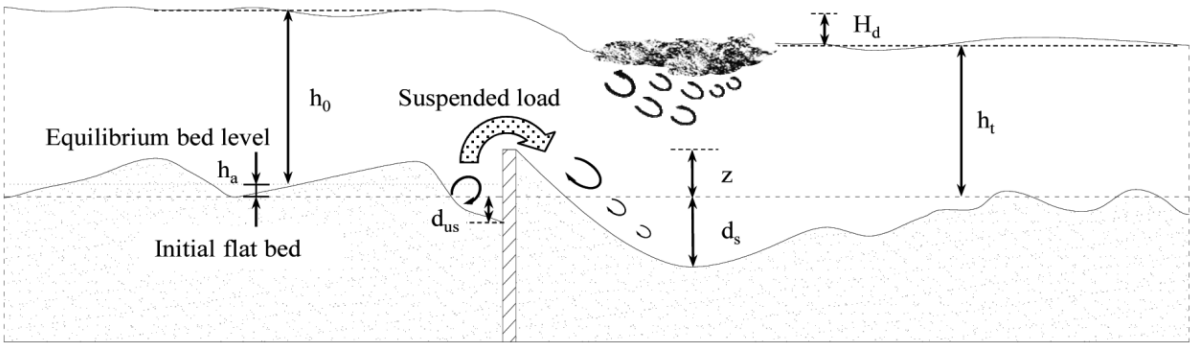


Figure 1. Illustration of a submerged weir's live bed scour [5].

https://doi.org/10.31436/iiumej.v26i3.3825

1.2. Study Area

The Kasa Low Weir in Pasir Mas, Kelantan, was constructed to regulate water levels for the pump stations that support paddy field irrigation. Completed on 19th April 2016 with a crest level of 1.8 m, the structure included adjustable submerged weirs, comprising two hydraulic tilting gates and a navigation lock for regulating water levels and facilitating boat passage. However, on 11th June 2018, two years after completion, a 130-meter section between the gates collapsed due to severe scouring, disrupting water regulation. A forensic investigation by the Department of Irrigation and Drainage (DID) in 2020 revealed extensive scour both upstream and downstream, particularly around the tilting gates, with depths ranging from -2 m to -14 m as depicted in Figure 2. The failure was attributed to the structure's inability to withstand hydrodynamic forces, causing seepage and scour. Due to ongoing instability, a Crump weir design as a fixed weir was one of the proposed designs for the upcoming rehabilitation, offering a more stable discharge coefficient to improve hydraulic performance and reduce the risk of future failure.

Figure 2. The area of the hydraulic tilting gates of Kasa Low Weir that has collapsed and washed away due to significant scouring.

1.3. Submerged Weir

Submerged weirs, including both adjustable and fixed types, are hydraulic structures installed below the water surface to regulate flow velocities, stabilise riverbeds, and manage water levels in rivers [6]. These weirs operate under two primary flow regimes: free flow, where downstream water levels remain below the crest, and submerged flow, where tailwater elevations exceed the crest, altering discharge characteristics. Scouring, the erosion of sediment around these structures, poses significant risks to their stability, particularly during high-flow events. For instance, live-bed scour conditions, where sediment transport occurs continuously, create dynamic bedforms that amplify scour depths downstream of submerged weirs [7-8].

Adjustable and fixed submerged weirs employ distinct mechanisms to mitigate scouring effects, striking a balance between hydraulic efficiency and structural resilience. While numerous studies have investigated hydraulic behaviour and associated scouring patterns downstream of submerged weirs, most have focused on either adjustable or fixed types in isolation, often through experimental setups with limited discharge scenarios. As a result, there is a lack of comprehensive, side-by-side numerical comparisons under identical flow conditions, particularly using computational fluid dynamics (CFD) modelling such as FLOW-3D [9]. This study presents a numerical comparison of sediment scour downstream of adjustable and fixed submerged weirs, which helps improve the understanding of both weir

https://doi.org/10.31436/iiumej.v26i3.3825

design and its influence on scouring, leading to better design and sediment management in rivers.

Submerged weirs are widely used in river engineering to regulate flow and manage sediment transport, but their performance in resisting downstream scouring varies depending on their design. Adjustable submerged weirs, such as hydraulic tilting gates, offer operational flexibility, while fixed weirs like the Crump type provide geometric stability. However, how these design differences influence flow behaviour and scouring under different flow conditions remains insufficiently explored. A direct comparison between the two is necessary better to understand their hydraulic performance and impact on bed erosion.

1.3.1. Adjustable Submerged Weir (Hydraulic Tilting Gate)

Adjustable submerged weirs, such as hydraulic tilting gates, offer operational flexibility to control water discharge and maintain desired upstream levels. The structural systems alter crest heights (± 2 mm precision) to optimise discharge distribution, reducing shear stresses during high-flow events [10-11]. These systems utilize rotational mechanisms around a horizontal axis to modulate discharge capacity, allowing for real-time adaptation to fluctuating hydrological conditions, such as flood events or sediment-laden flows. However, this adjustability can lead to varying flow regimes, which in turn influence the development and extent of cyclical scour patterns downstream. Past studies reveal that adjustable weirs under live-bed conditions experience oscillating scour depths (up to 44% variation) due to alternating jet intrusion and deposition phases [12]. Experimental analyses using particle image velocimetry (PIV) reveal that turbulence intensities near tilting gates reach 2.5 times the upstream values, creating sediment entrainment pulses synchronised with the gate opening [13]. The dynamic nature of flow over these gates often results in complex turbulence and sediment transport processes, necessitating detailed analysis to predict scour characteristics downstream of the structure.

1.3.2. Fixed Submerged Weir (Crump Weir)

Fixed submerged weirs, such as the Crump weir, rely on geometric stability to induce predictable scour patterns. This structure, characterised by triangular profiles with upstream slopes of 1:2 and downstream slopes of 1:5, maintains predictable performance under modular flow [14]. Under modular flow conditions, Crump weir design maintains a discharge coefficient (Cd) of less than 5% across a 10:1 head range, ensuring reliable performance in a sediment-laden environment [15]. The weir's sloped faces create a cascading energy dissipation mechanism that reduces downstream maximum shear stresses by 54% compared to vertical-faced structures when the tailwater height exceeds 1.2 times the critical depth. The experimental analyses reveals that Crump weirs achieve 23% greater energy dissipation efficiency than tilting gates under equivalent hydraulic loads due to their established crest geometry and optimised slope ratios [16]. The fixed profile of the Crump weir minimises operational complexity and mechanical maintenance, making the Crump weir indispensable for long-term infrastructure resilience against scouring risks.

1.4. FLOW-3D

FLOW-3D employs a finite volume method to solve the three-dimensional, incompressible Reynolds-averaged Navier-Stokes (RANS) equations, coupled with the Volume of Fluid (VOF) technique for tracking free surfaces [17]. The solver integrates the continuity and momentum conservative equations, formulated as Eq. (1) and Eq. (2):

Continuity equation:

https://doi.org/10.31436/iiumej.v26i3.3825

$$\frac{\partial u_i}{\partial x_i} = 0 \tag{1}$$

Momentum equation:

$$\frac{\partial u_i}{\partial \mathrm{t}} + u_j \frac{\partial u_i}{\partial x_j} = -\frac{1}{\rho}\frac{\partial \mathrm{p}}{\partial x_i} + \nu \frac{\partial^2 u_i}{\partial x_j^2} + \frac{\partial \tau_{ij}}{\partial x_j} \tag{2}$$

where τ_{ij} represents Reynolds stresses, which are modelled using turbulence closure schemes. The Fractional Area-Volume Obstacle Representation (FAVOR^{TM}) method is utilised to embed complex geometries into the Cartesian grid, ensuring accurate boundary conditions while minimising mesh generation complexity.

1.4.1. Turbulence Model

The RNG turbulence model was employed for this study for several compelling reasons. First, it is particularly well-suited for simulating turbulent flow over a weir, which is critical for accurately capturing the complex hydrodynamic interactions involved in scour processes. Second, among the turbulence models available in commercial computational fluid dynamics software, the RNG model is recognised for its superior accuracy and robustness in predicting flow behaviour relevant to scouring. Lastly, due to its enhanced capability to account for high levels of turbulence generated by fluid flow through control structures, the RNG model often demonstrates improved performance in scour simulations. The listed attributes collectively justify its selection for the present analysis [18]. The transport equations for turbulent kinetic energy (k) and dissipation rate (ϵ) are according to Eq. (3) and Eq. (4):

$$\frac{\partial(\mathrm{pk})}{\partial \mathrm{t}} + \frac{\partial(\mathrm{pku_j})}{\partial \mathrm{x_j}} = \frac{\partial}{\partial \mathrm{x_j}}\left(\alpha_\mathrm{k}\mu_\mathrm{eff}\frac{\partial \mathrm{k}}{\partial \mathrm{x_j}}\right) + \mathrm{Gk} - \mathrm{p}\epsilon \tag{3}$$

$$\frac{\partial(\mathrm{p}\epsilon)}{\partial \mathrm{t}} + \frac{\partial(\mathrm{p}\epsilon \mathrm{u_j})}{\partial \mathrm{x_j}} = \frac{\partial}{\partial \mathrm{x_j}}\left(\alpha_\mathrm{k}\mu_\mathrm{eff}\frac{\partial \epsilon}{\partial \mathrm{x_j}}\right) + \mathrm{C_{1\epsilon}}\frac{\epsilon}{\mathrm{K}}\mathrm{G_k} - \mathrm{C_{2\epsilon}}\rho\frac{\epsilon^2}{\mathrm{K}} - \mathrm{R_\epsilon} \tag{4}$$

Here, $\mu_{eff} = \mu+\mu_t$, with $\mu_t = \rho C_\mu k^2/\epsilon$. The RNG model introduces the Rϵ term to address rapid strain and streamline curvature effects, enhancing accuracy in flows with high shear or rotation. Unlike the standard k-ε model, the RNG variant dynamically adjusts Cμ (turbulent viscosity coefficient) and employs an analytically derived turbulent Prandtl number (α_k, α_ϵ), reducing reliance on empirical constants.

1.4.2. Sediment Scour Model

The prediction of sediment scouring downstream of weirs is a critical concern in hydraulic engineering, as scouring can undermine the structural stability of these hydraulic control structures. The Meyer-Peter and Müller (MPM) equation is frequently employed within FLOW-3D to simulate bedload sediment transport and scour evolution in such contexts, relating sediment flux to excess shear stress in Eq. (5) [19]:

$$\Phi = 8(\theta - \theta_{\mathrm{cr}})^{1.5} \tag{5}$$

where $\Phi = q_b/\sqrt{g d_{50}^3}$ is the dimensionless bedload transport rate, $\theta = \tau/(\rho g d_{50})$ is the Shields parameter, and θ_{cr} is the critical Shields parameter for incipient motion. The bed shear stress τ is computed from the RNG-resolved velocity field, while θ_{cr} is calibrated using sediment grain size and density [20].

Morphological evolution of the bed is tracked through the Exner equation, as expressed in Eq. (6):

https://doi.org/10.31436/iiumej.v26i3.3825

$$(1 - \lambda_p)\frac{\partial \eta}{\partial t} + \frac{\partial q_b}{\partial x} = 0 \quad (6)$$

where η is bed elevation and λ_p is sediment porosity. FLOW-3D enhances the MPM formulation by incorporating 3-dimensional effects, such as secondary currents and turbulent bursting, which amplify local scour by up to 30% compared to depth-averaged approaches. Scour depth predictions around bridge piers show less than 10% error when validated against experimental data using RNG-resolved turbulence [21].

2. METHODOLOGY

2.1. Model Validation

To assess the depth-averaged velocity along the Kelantan River, Acoustic Doppler Current Profiler (ADCP) instruments were strategically deployed at three key locations: upstream of Sultan Yahya Petra Bridge (Q1/WL1), downstream of Kasa Low Weir (Q2/WL2), and upstream of Sheet Pile Kemumbu structure (Q3/WL3) as presented in Figure 3. Among these, station Q2/WL2, positioned immediately downstream of the Kasa Low Weir, was selected for validation analysis on the FLOW-3D model due to its critical hydrodynamic behaviour influenced by the weir structure. For depth-averaged velocity, data were collected from station Q2, located downstream of Kasa Low Weir, from 13th October 2024 at 17:00 until 2nd November 2024 at 15:40.

Figure 3. Location of ADCP deployed at the Kelantan River.

The validation model was developed based on the actual 1 km upstream and downstream surveyed bathymetry and cross-sectional topography of the Kelantan River at Kasa Low Weir to represent the physical conditions of the study area accurately. In FLOW-3D, the bathymetry was modelled as a fixed bed to focus solely on hydrodynamic analysis, without the influence of sediment transport or bed deformation. This approach allowed for a more precise comparison between the simulated and measured flow characteristics. Validation was performed using field data collected from the ADCP at station Q2/WL2, which is located approximately 1.11 km downstream of the Kasa Low Weir, as shown in Figure 4.

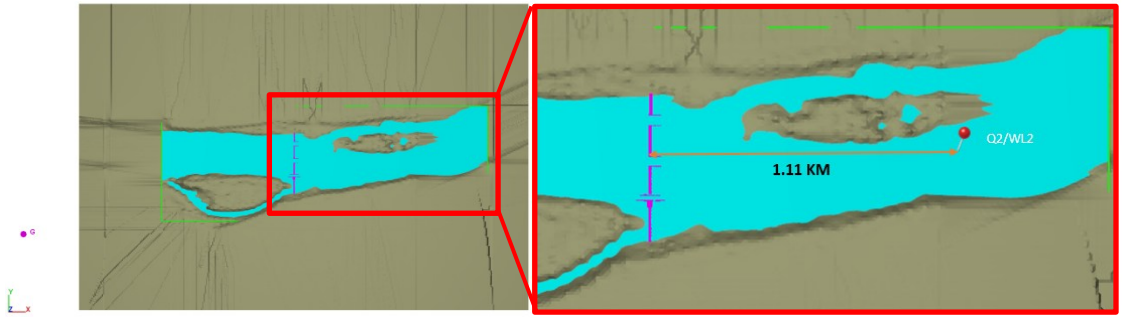


Figure 4. Model setup for validation in FLOW-3D

https://doi.org/10.31436/iiumej.v26i3.3825

To estimate the upstream discharge conditions at the Kasa Low Weir, measurements were carried out using an Acoustic Doppler Current Profiler (ADCP) at Station Q3/WL3, located upstream near the Sheet Pile Kemumbu. On 2 November 2022, depth-averaged velocity and were recorded at 10-minute intervals over one hour from 14:40 to 15:40. During this time, the average current velocity ranged from 0.788 m/s to 0.819 m/s, with an overall average of 0.816 m/s. To ensure accurate discharge calculation, both ADCP probes, Q2 (downstream of the weir) and Q3 (upstream), were deployed in a coordinated way. According to Figure 5, Station Q2 was deployed on 13 October 2022 at 17:50, while Station Q3 was deployed on 15 October 2022 at 11:00. Q2 was retrieved on 2 November 2022 at 15:40, and Q3 on 10 November 2022 at 09:50. Most importantly, both stations collected data at the same time on 2 November 2022 between 14:40 and 15:40. This synchronised measurement ensures consistency when comparing flow conditions upstream and downstream of the weir and improves the accuracy of model validation and discharge estimation.

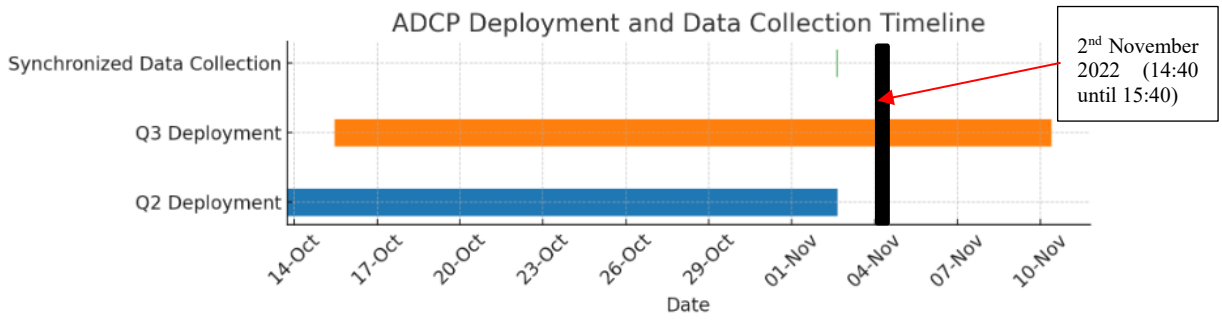


Figure 5. Timeline of ADCP deployments and synchronised data collection.

The upstream discharge was calculated using the area–velocity method. The river cross-section was divided into several segments, each with a measured width of 62.5 m and varying depths from 0.8 m to 6.0 m. This gave a total cross-sectional area of 962.5 m². By multiplying this area by the average velocity of 0.816 m/s, the estimated upstream river discharge was 785.4 m³/s. This value represents typical flow conditions upstream of the Kasa Low Weir and serves as a key input for hydrodynamic validation of the model in Figure 6.

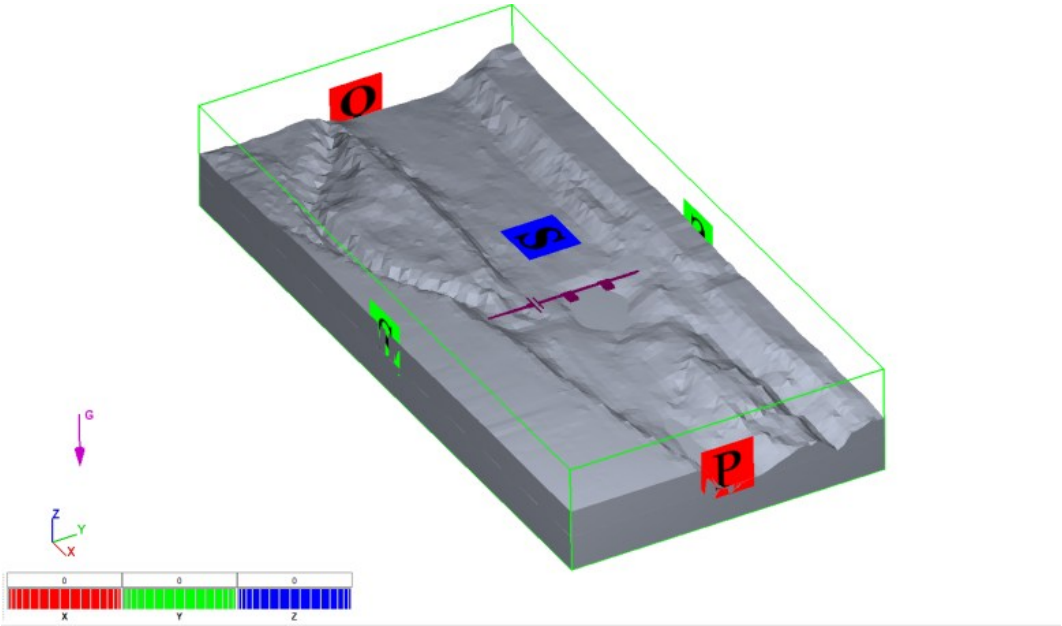


Figure 6. Boundary conditions for model validation.

https://doi.org/10.31436/iiumej.v26i3.3825

The validation of the numerical model will be carried out by comparing the simulated results with observed data obtained from the ADCP. Two different mesh sizes, 1.5 m (finer mesh) and 3.0 m (coarser mesh), will be tested in the model to evaluate the impact of mesh resolution on simulation accuracy, as shown in Figure 7. The comparison will focus on key flow parameters such as depth-averaged velocity at the location of the ADCP probe.

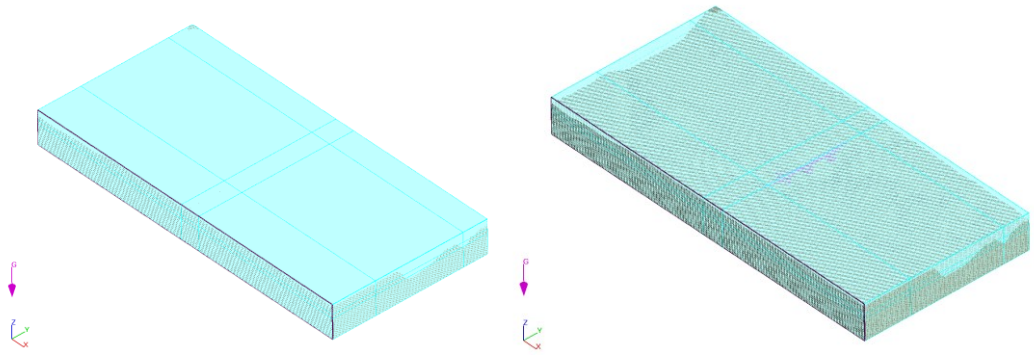


Figure 7. Comparison of validation results using two different mesh sizes (1.5 m and 3.0 m) in the model.

To quantify the agreement between the simulated and observed flow velocity, the mean absolute error (MAE), mean squared error (MSE), and coefficient of determination (R^2) were calculated, as presented in Table 1. The RMSE, MAE, and R^2 are calculated for the data set. Root Mean Squared Error (RMSE) in Eq. (7) serves as a critical measure by quantifying the average magnitude of prediction errors. RMSE computes the square root of the mean squared differences between predicted and observed values, thereby penalising larger deviations more heavily. This characteristic makes RMSE particularly sensitive to significant errors, offering a comprehensive assessment of model accuracy in the same units as the output variable [22].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2} \tag{7}$$

In Eq. (8), the Mean Absolute Error (MAE) provides an alternative validation metric that measures the average absolute deviations between predicted and true values without biasing larger errors. By treating all errors uniformly, MAE conveys an interpretable and straightforward metric of model performance, reflecting the expected magnitude of errors in predictions regardless of their direction or scale.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}i| \tag{8}$$

The Coefficient of Determination (R^2), as shown in Eq. (9), serves as a statistical index that expresses the proportion of variance in the dependent variable accounted for by the numerical model. R^2 values range from 0 to 1, where higher values indicate stronger explanatory power and a better fit. This metric is essential in validating model effectiveness by quantifying how well the model captures the variability inherent in the observed data, thereby indicating its predictive reliability [23].

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{9}$$

From the equations, the symbol y_i refers to the actual observed value at the i-th data point, which represents the true measurement obtained from field data. $\hat{y}_i$ denotes the predicted value generated by the model at the same i-th point. These two values are compared to assess how closely the model simulates real conditions. The variable n represents the total number of

https://doi.org/10.31436/iiumej.v26i3.3825

observations or data points used in the comparison. Additionally, in the calculation of the coefficient of determination R^2, the mean of all observed values, represented as $\bar{y}$, is used to quantify how much of the variance in the observed data is explained by the model.

Table 1. Comparison of observed and simulated depth-averaged velocity at Station Q2/WL2 for mesh sizes of 1.5 m and 3.0 m over time.

Time step (s)	Observed Data	Simulated Data	
	Station Q2/WL2	1.5 m	3.0 m
0	0.747	0.760	0.740
600	0.754	0.742	0.754
1200	0.772	0.771	0.771
1800	0.681	0.678	0.681
2400	0.719	0.717	0.729
3000	0.135	0.146	0.250
3600	0.101	0.144	0.007

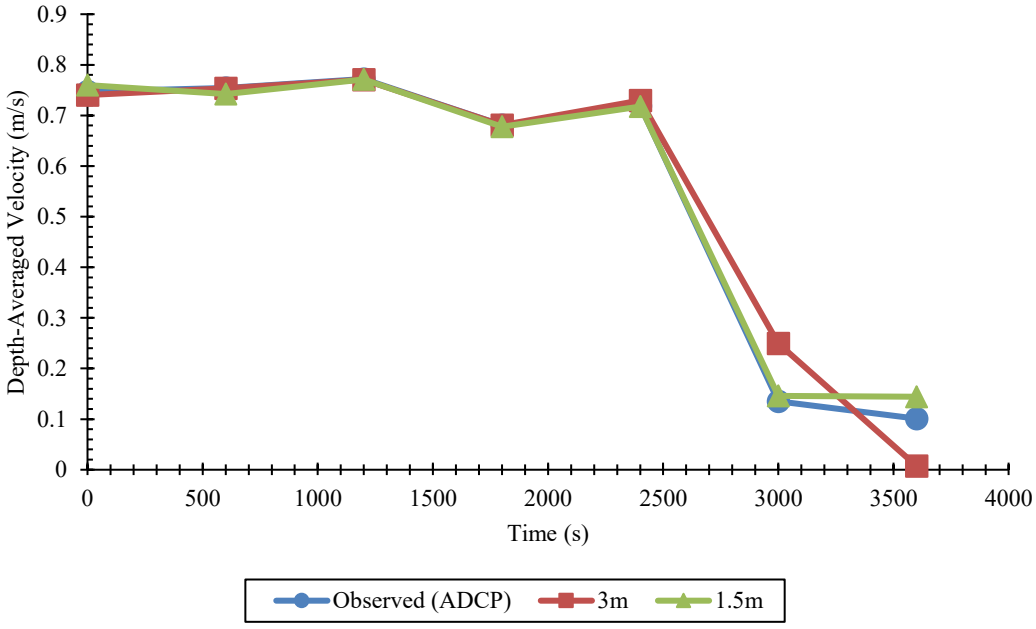


Figure 8. Depth-averaged velocity variation over time from observations and simulations using 1.5 m and 3.0 m mesh sizes.

Table 2: Statistical comparison between simulated and observed flow velocity for different mesh sizes.

Mesh sizes (m)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	Coefficient of Determination (R^2)
1.5 (finer mesh)	0.018	0.012	0.998
3.0 (coarser mesh)	0.056	0.033	0.961

The comparative analysis between simulated depth-averaged velocity results and the observed ADCP data is presented in both Table 1 and Figure 8. The results show in Table 2 that the finer mesh (1.5 m) closely follows the observed trend across all time steps, with lower error metrics: an RMSE of 0.018 and an MAE of 0.012. In contrast, the coarser mesh (3.0 m) yields a higher RMSE of 0.056 and MAE of 0.033, indicating a greater deviation from the

https://doi.org/10.31436/iiumej.v26i3.3825

observed values. This is further supported by the coefficient of determination (R^2), where the finer mesh achieved 0.998, suggesting a near-perfect fit, while the coarser mesh produced a slightly lower R^2 of 0.961. The performance difference is particularly evident at later steps, at 3000 s until 3600 s, where the 3.0 m mesh significantly underestimates the depth-averaged velocity. Overall, the results demonstrate that finer mesh resolution enhances the model's accuracy in replicating the observed flow behaviour. The results show that the simulated water levels closely match the measured values. While slight deviations were observed at certain points, the overall simulation demonstrates reliable performance. Given its higher accuracy, the 1.5 m mesh size will be selected for the subsequent sediment transport modelling.

2.2. Sediment Transport Model

To focus specifically on sediment transport processes near the structure and to reduce computational time, the full-scale model domain was trimmed to include only the region surrounding the weir. This reduced boundary setup allowed for a more detailed analysis of local scour and sediment dynamics while minimising unnecessary calculations in areas not critical to the study. The refined domain was then imported into FLOW-3D for sediment transport simulation.

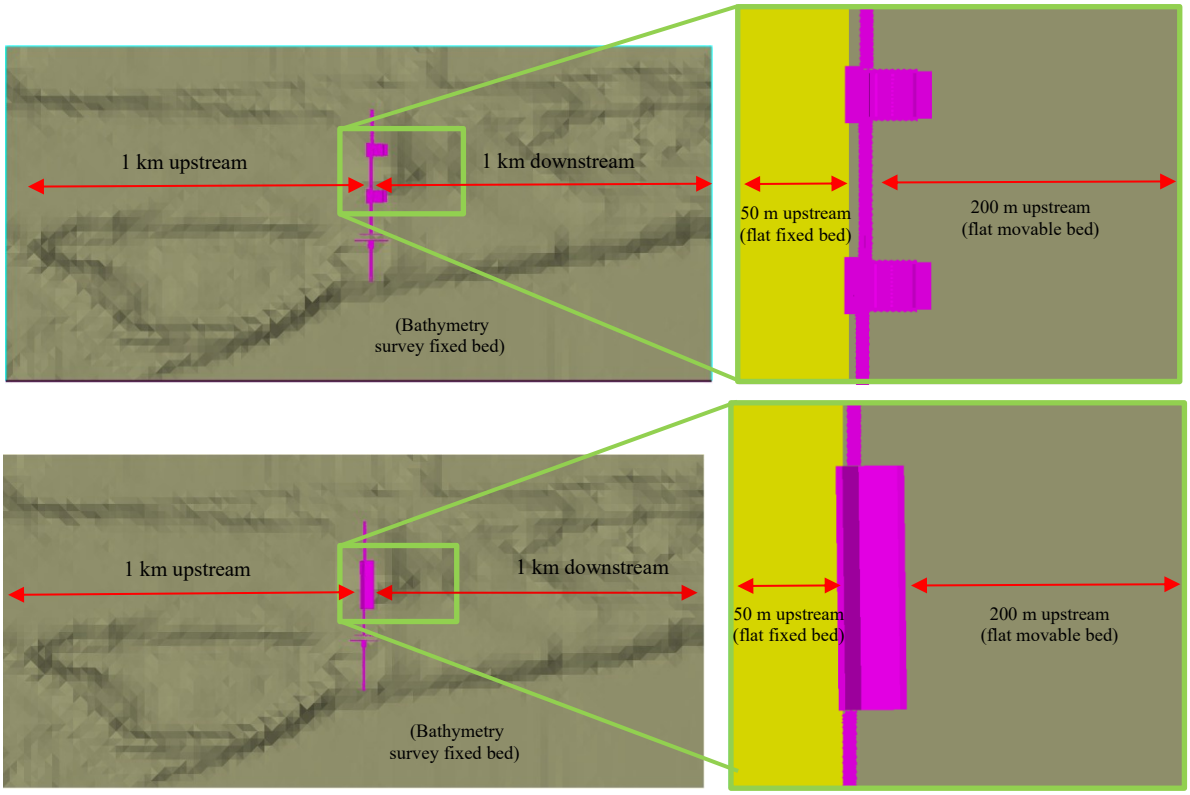


Figure 9. Reduced boundary model around weir structures used for sediment transport simulation.

2.2.1. Geometry Setup

In this study, a comparison is made between an adjustable (two hydraulic tilting gates), which allows active flow control through gate opening for variable flow conditions, and a fixed (Crump weir) submerged weir, a simpler and passive structure that permits overflow based on upstream water levels, as shown in Figure 10(a) and Figure 10(b). In Figure 11(a) and Figure 11(b), the geometry design of the hydraulic tilting gates is based on a real-life project used as

https://doi.org/10.31436/iiumej.v26i3.3825

a case study, while the Crump weir is considered as an alternative to replace the gate. Both geometric models are created using SketchUp software.

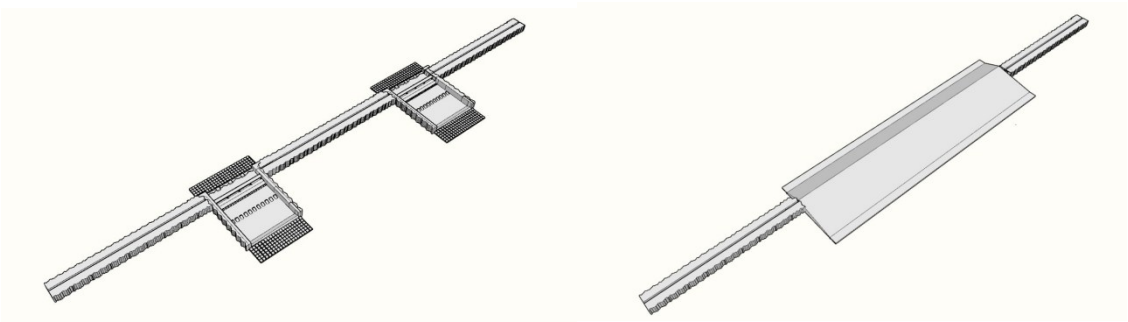
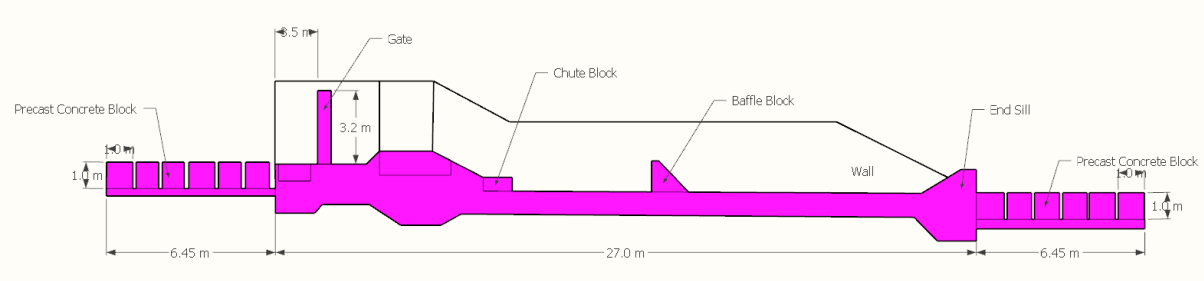
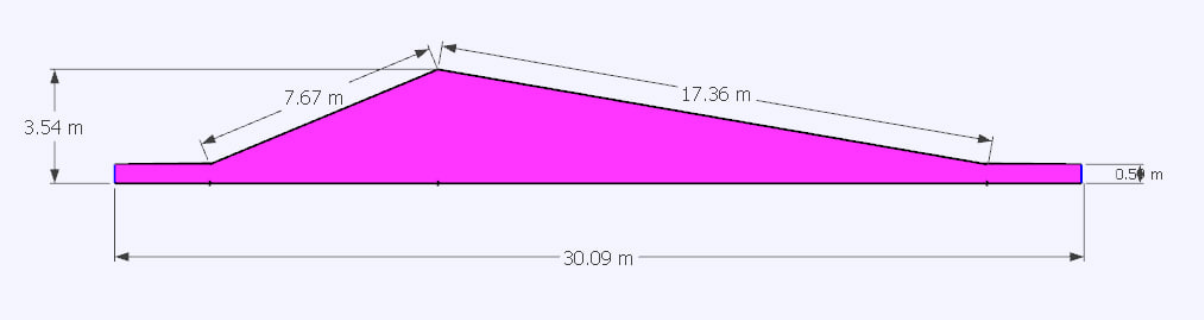


Figure 10. Dimension of (a) hydraulic tilting gate and (b) Crump weir in SketchUp software in isometric view.



(a)



(b)

Figure 11. Dimension of (a) hydraulic tilting gate and (b) Crump weir from side view.

2.2.2. Boundary Setup

In FLOW-3D, specifying the simulation time, physics, fluid characteristics, meshing, boundary conditions, and initial conditions are important parameters for modelling riverbed scouring. The simulation runs for 3600 seconds to accurately capture flow dynamics downstream of both weir structures. The simulation setup includes Shallow Water Approximation, Gravity (Z = -9.81 m/s^2), Turbulence and Viscosity (Renormalised group (RNG) model), and Density (evaluated as a function of other parameters, such as temperature or scalars). To ensure alignment with the actual real-life project water levels, the initial condition sets the fluid elevation at +1.70 m from datum for both weir configurations, as shown in Figure 12(a) and Figure 12(b).

https://doi.org/10.31436/iiumej.v26i3.3825

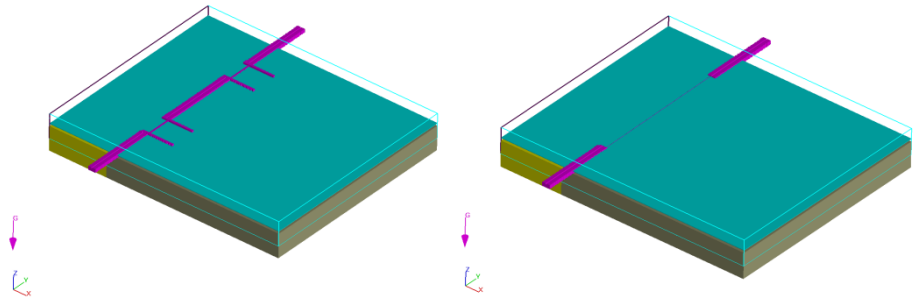


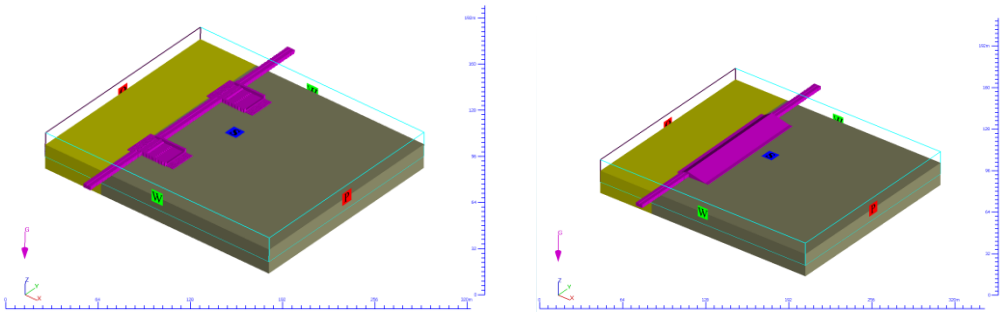
Figure 12. Initial conditions for (a) adjustable submerged weir (hydraulic tilting gate) and (b) fixed submerged weir (Crump weir) model setup configurations.

The summary of the Sediment Transport function for both weir configurations (Meyer-Peter & Müller equation) is as follows, based on the conditions of the Kelantan riverbed as presented in Table 3.

Table 3. Overview of the Sediment Transport Physics in FLOW-3D in this study.

Name	Diameter [m]	Density [kg/m^3]	Critical Shields Number	Entrainment Coefficient	Bed Load Coefficient	Maximum Packing Fraction
Sediment 1	0.0074	2650	0.05	0.018	12	0.0001
Sediment 2	0.0070	2650	0.05	0.018	12	0.0001
Sediment 3	0.0085	2650	0.05	0.018	12	0.0001

Several constraints must be established in FLOW-3D modelling. FLOW-3D imports the geometric model in STL format. As seen in both Figure 13(a) and 13(b), a rectangular flume with 170 m length, 200 m width, and a height of 20 m using the mesh barrier. The dimensions provide enough space for both upstream and downstream hydraulic structures to observe flow behaviour and scour development accurately. The upstream bed was defined as a fixed (solid) bed to maintain stable flow conditions before reaching the weir structure. The downstream bed was modelled as a movable bed, allowing sediment transport and deformation to occur. This configuration was chosen to focus on observing and analysing local scour development downstream of adjustable and fixed submerged weirs under varying flow conditions.



(a) (b)

Figure 13. Boundary setup for both models.

The boundary conditions include Wall (W) for non-slip boundary surfaces, Symmetry (S) to reduce computational load, Specified Pressure at the outlet to control hydrostatic conditions, and Volume Flow Rate (Q) to regulate inflow, representing a range of hydraulic loading conditions. In Table 4, a total of 20 simulations were performed for each structure type, with

https://doi.org/10.31436/iiumej.v26i3.3825

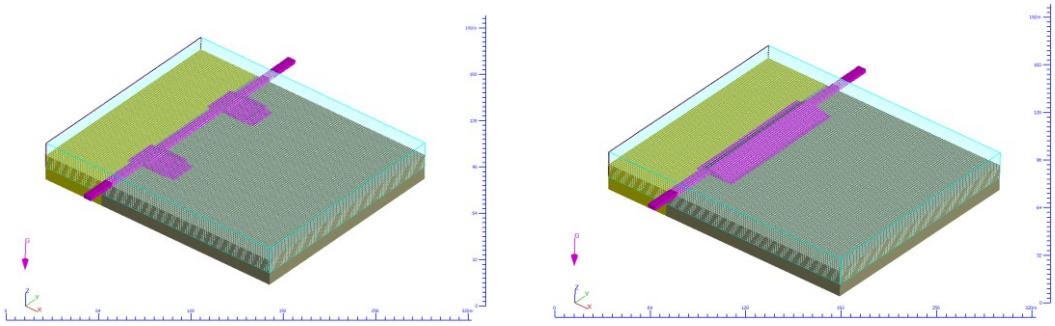
discharge values incrementally increasing from 182.30 m^3/s to 3666.00 m^3/s under a 2-year ARI of the Kelantan River using a percentile-based approach. This range was selected to capture the effects of low to high flow scenarios on the development of local scour downstream of the weirs. The classification into low, moderate, and high discharges is justified using a percentile-based approach [24]. By gradually increasing the flow rate in each simulation, this study was able to observe how different discharge levels affect flow velocity, shear stress, and the pattern of bed erosion around the hydraulic tilting gate and Crump weir.

Table 4: Sediment Transport simulation runs for both the adjustable and fixed submerged weir in this study.

Run(s)	Discharge, Q (m^3/s): Low	Run(s)	Discharge, Q (m^3/s): Moderate	Run(s)	Discharge, Q (m^3/s): High
1	183.30	**8**	1466.40	**15**	2749.50
2	366.60	**9**	1649.70	**16**	2932.80
3	549.90	**10**	1833.00	**17**	3116.10
4	733.20	**11**	2016.30	**18**	3299.40
5	916.50	**12**	2199.60	**19**	3482.70
6	1099.80	**13**	2382.90	**20**	3666.00
7	1283.10	**14**	2566.20		

2.2.3. Meshing Setup

In both Figure 14(a) and Figure 14(b), a structured mesh was generated to ensure accurate and stable numerical simulations in computational fluid dynamics (CFD) analysis [25]. The mesh block was defined as uniform cell sizes of 1.5 meters in the X, Y, and Z directions, resulting in a total of 195,377 cells. The grid was distributed across the domain with 133 cells along the X-axis, 113 cells along the Y-axis, and 13 cells along the Z-axis. To assess mesh quality, key indicators such as the maximum adjacent cell size ratio and maximum aspect ratio were evaluated in Table 5. The maximum adjacent ratios for X, Y, and Z directions were 1.00001, 1.00001, and 1.0000. Similarly, the maximum aspect ratios for each plane (X-Y, Y-Z, and Z-X) were 1.00045, 1.02263, and 1.02308, respectively. To ensure mesh quality, the quantity and spacing of cells in the x, y, and z directions must be configured such that the maximum adjacent cell size ratio remains below 1.25, and the maximum aspect ratio does not exceed 3.0 [26]. These values confirm that the mesh used in the simulation is of high quality, ensuring numerical stability and reliable results for analysing flow and scouring behaviour around the hydraulic structures.



(a) (b)

Figure 14. Meshing setup for both models.

https://doi.org/10.31436/iiumej.v26i3.3825

Table 5. Types and information of the mesh.

Maximum adjacent cell size ratio	Maximum aspect ratio
X direction: 1.00001 < 1.25	X-Y direction: 1.00045 < 3.0
Y direction: 1.00001 < 1.25	Y-Z direction: 1.02263 < 3.0
Z direction: 1.00000 < 1.25	Z-X direction: 1.02308 < 3.0

3. RESULT AND DISCUSSION

3.1. Velocity at Bed

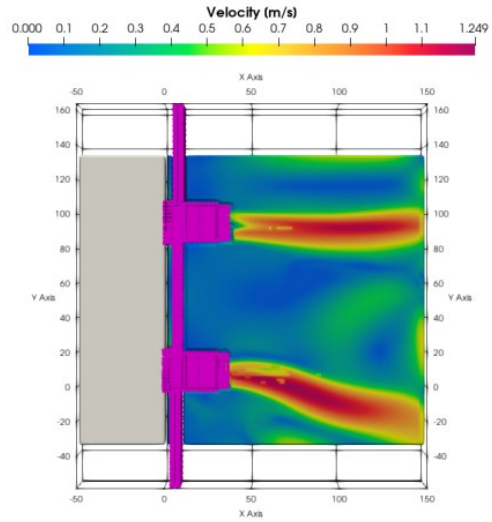
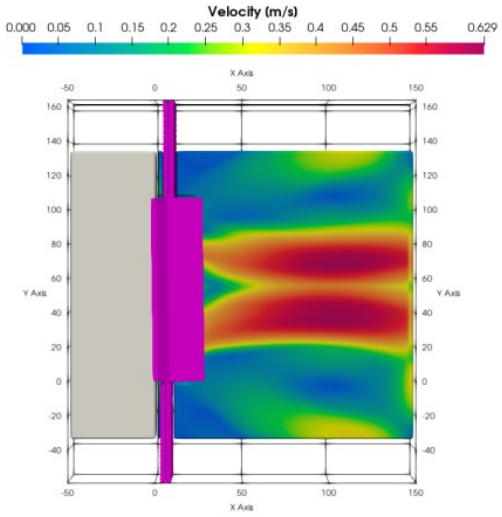
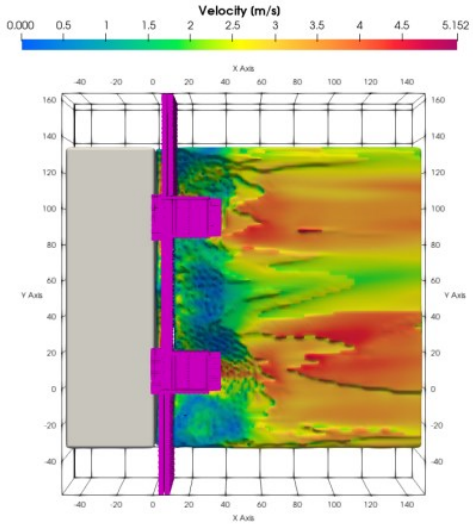
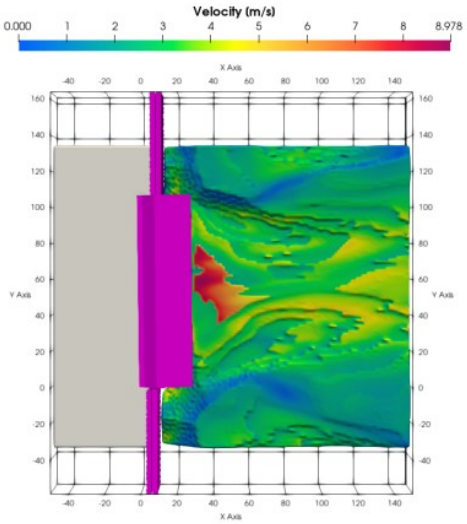
Understanding velocity near the riverbed is important for assessing how sediment erosion occurs downstream of hydraulic structures. In submerged weirs, the bed-level velocity—where water interacts with sediment—greatly influences the size and shape of scour holes. This velocity is affected by weir design, flow rate, and sediment behaviour. High-velocity areas shown in yellow, orange, and red colour contour range indicate strong shear forces that can move sediment and create deeper scour. In contrast, low-velocity areas (shown in blue and green) suggest low-energy zones where sediment tends to settle.

This section compares the velocity patterns at scour depth for two weir types: the adjustable submerged weir (hydraulic tilting gate) and the fixed submerged weir (Crump weir). From the 20 simulations, Run 1 and Run 20 were selected in Table 6 to show how minimum and maximum flow rates impact downstream velocity and scour development. These runs represent a range of flow conditions, from low to extreme, and highlight how different weir designs influence flow behaviour and scouring patterns.

Under low-flow conditions, represented by Run 1 (Q = 183.30 m^3/s), the velocity at the scoured depth reaches up to 1.249 m/s for the Adjustable Submerged Weir, while the Fixed Weir shows a much lower peak of around 0.629 m/s. The velocity contours for both structures are mostly in the blue and green range, indicating low flow energy and a minimal risk of sediment movement. However, the adjustable weir still generates a more concentrated flow jet, suggesting that it may initiate localised erosion even under low discharge conditions, unlike the fixed weir. Under the extreme flow condition of Run 20, the Fixed Weir shows a clear dominance in velocity, reaching 8.978 m/s, while the Adjustable Weir peaks at only 5.152 m/s. The fixed structure generates a highly energetic flow field, with visible signs of turbulence. In contrast, the adjustable weir's jet appears more stable but less intense, possibly due to flow limitations at the gate opening. Overall, the adjustable submerged weir tends to concentrate flow into high-velocity jets, which can cause deeper scour. In contrast, the fixed submerged weir tends to spread the flow energy more broadly, leading to wider and potentially more uniform erosion. These differences highlight how the geometry and design of submerged weirs significantly affect flow behaviour and scour characteristics at the sediment bed.

https://doi.org/10.31436/iiumej.v26i3.3825

Table 6. Comparison of Velocity at Scour Depths Between Original (Adjustable Submerged) Weir and Crump (Fixed Submerged) Weir Across Simulated Flow Run 1 and 20.

Run	Adjustable Submerged Weir (Hydraulic Tilting Gate)	Fixed Submerged Weir (Crump Weir)
1 (Q = 183.30 m^3/s)	 Maximum: 1.249 m/s	 Maximum: 0.629 m/s
20 (Q = 3666.00 m^3/s)	 Maximum: 5.152 m/s	 Maximum: 8.978 m/s

The following graph in Figure 15 shows the relationship between velocity at scour depth (V) and discharge (Q) across all 20 simulated runs. On average, the adjustable submerged weir records slightly higher velocity at scour depth, with an average of 4.848 m/s compared to the fixed submerged weir. This trend is supported by the mean values, where the adjustable submerged weir shows 4.590 m/s, while the fixed submerged weir records 4.300 m/s. Although the difference may appear modest numerically, the implications are significant in terms of sediment mobilisation, as velocity increases can result in exponential rises in shear stress and erosion potential.

The maximum velocity for a fixed submerged weir is slightly higher than that of an adjustable submerged weir, with values of 8.978 m/s and 5.152 m/s, respectively. This suggests that under the highest flow conditions, a fixed submerged weir can produce flow intensities equal to or even greater than an adjustable submerged weir. However, the minimum recorded

https://doi.org/10.31436/iiumej.v26i3.3825

velocity is substantially lower for the fixed submerged weir (0.629 m/s) than the adjustable submerged weir (1.249 m/s), indicating more variability and potentially more extensive zones of low-energy flow, which can favour sediment deposition rather than scouring. The velocity at scour depth comparison reveals that an adjustable submerged weir generally produces higher flow velocities during low to medium discharge conditions, which may trigger earlier localised scour. In contrast, at higher discharge levels (after Run 12), the fixed submerged weir begins to match or exceed the adjustable submerged weir in velocity, indicating a shift in hydraulic behaviour. This reflects the fixed submerged weir's tendency to disperse flow energy more broadly, potentially leading to wider yet less concentrated zones. These findings highlight how different weir designs influence the extent of sediment erosion, reinforcing the importance of detailed velocity analysis for effective sediment control.

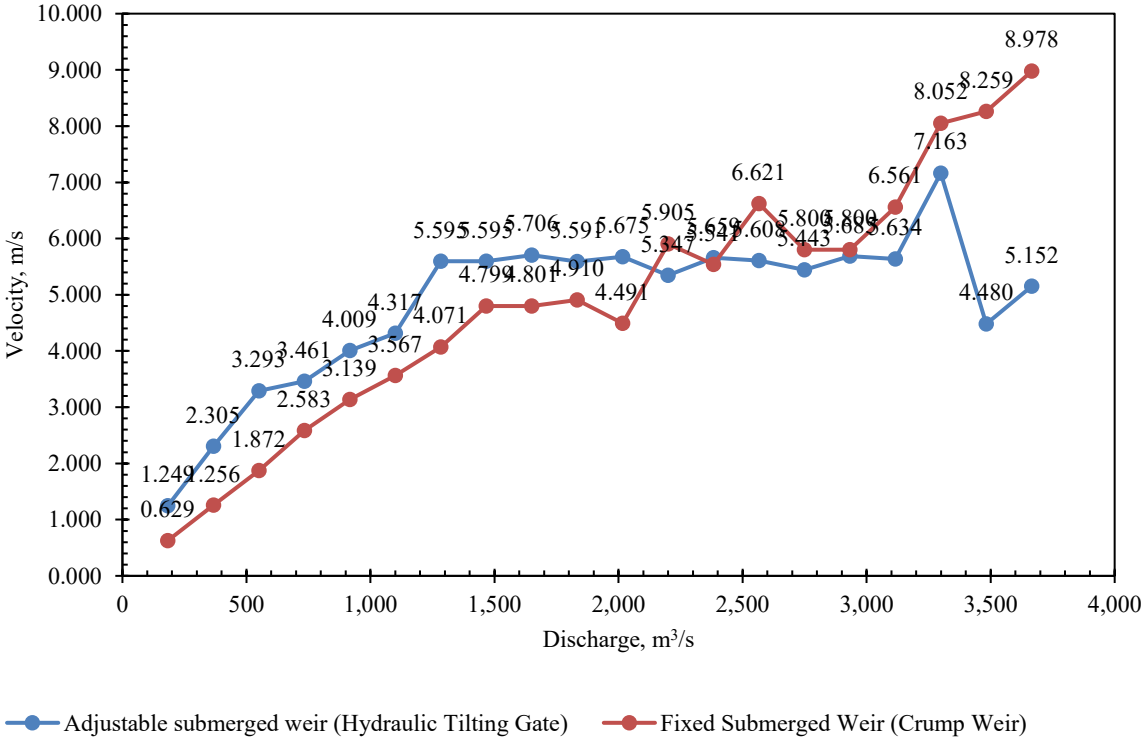


Figure 15. Velocity at Bed (V) versus Discharge (Q) for Adjustable Submerged Weir (Hydraulic Tilting Gate) and Fixed Submerged Weir (Crump Weir).

At higher discharges, the bed velocities downstream of the fixed Crump weir are higher compared to adjustable tilting gates, mainly due to the differences in their shape and flow behaviour. The sloping face of the Crump weir guides higher velocity water directly to the bed, allowing more energy to be transferred and creating stronger forces near the bottom. Conversely, adjustable hydraulic tilting gates produce a jet that is lifted above the bed, and much of its energy is lost through turbulence and recirculation before it reaches the bottom. Tailwater conditions also play a role, as the Crump weir often creates a steeper water-surface slope, which increases flow speed, while gates tend to raise downstream water levels and reduce velocity at the bed. As a result, the Crump weir sustains stronger near-bed velocities, which can influence the extent and pattern of scouring downstream.

3.3. Scour Depth

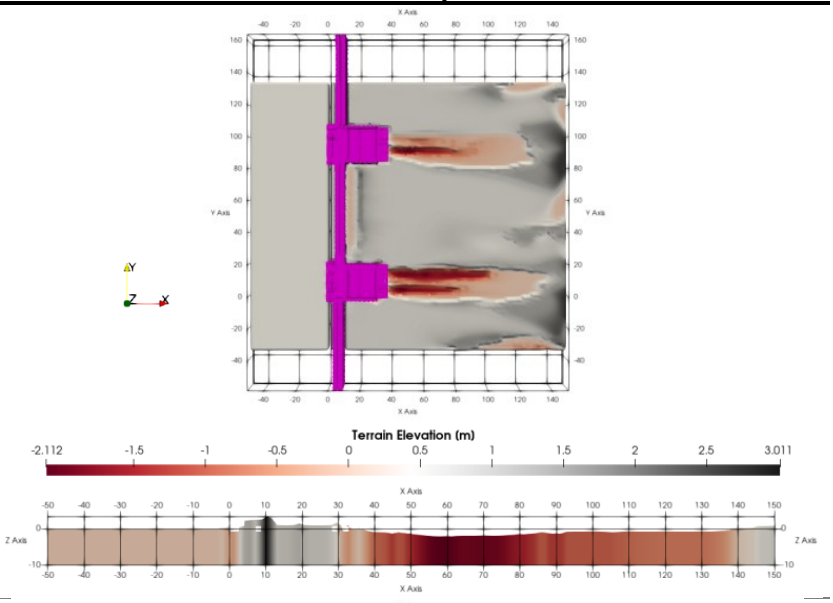
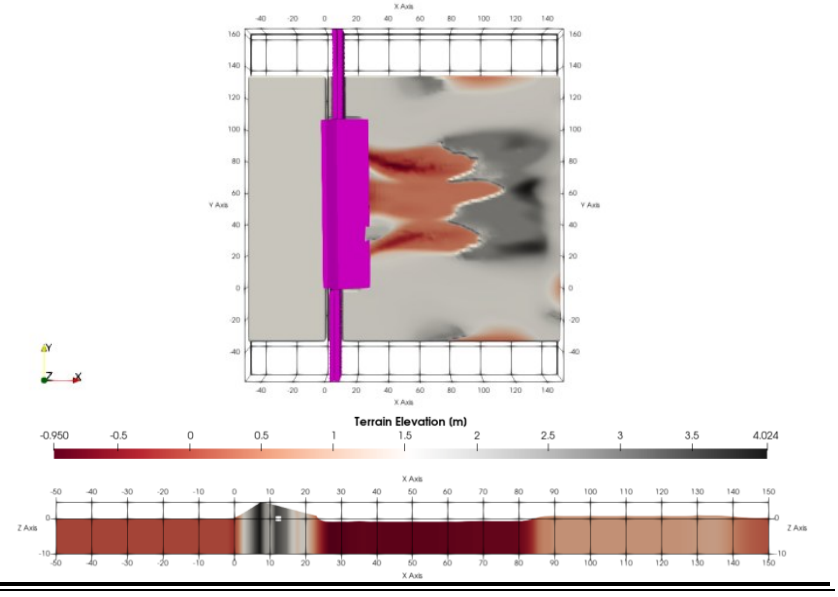
To study how weir type affects scour development, terrain elevation data from FLOW-3D simulations were analysed for 20 discharge runs, from low to high flow. Each simulation run

https://doi.org/10.31436/iiumej.v26i3.3825

provides a three-dimensional representation of the bed morphology, from which the scour depth is quantified by identifying the difference between the initial and final terrain elevations. The grey area contour plot shows deposition, while the red area contour plot indicates scouring. Table 4 below shows one example of Run 5 at Q = 916.50 m^3/s, comparing the effects of an adjustable submerged weir (Hydraulic Tilting Gate) and a fixed submerged weir (Crump Weir).

For this run, the adjustable submerged weir shows deeper and concentrated scour just downstream of the structure, shown as dark red areas. This is likely due to strong, high-velocity flow jets from the gate. In contrast, the fixed Crump weir results in a wider, more evenly spread erosion pattern, with less depth, suggesting more diffused flow energy. The same analysis method was used for all 20 runs, with constant simulation settings to ensure fair comparison. Overall, the adjustable weir caused stronger and more focused scouring, especially at higher flows, while the fixed weir led to broader but shallower erosion. These results highlight how the design of a weir affects erosion patterns and sediment movement in open channels.

Table 6. Terrain elevation comparison from FLOW-3D for adjustable (Hydraulic Tilting Gate) and fixed (Crump Weir) submerged weirs at Q = 916.50 m^3/s.

Run	Type of weir	Scour Depth
5 (Q = 916.50 m^3/s)	Adjustable Submerged Weir (Hydraulic Tilting Gate)	
	Fixed Submerged Weir (Crump Weir)	

https://doi.org/10.31436/iiumej.v26i3.3825

Figure 16 illustrates the relationship between scour depth (Ds) and discharge (Q) across 20 simulated runs for two types of weirs: an adjustable submerged weir (Hydraulic Tilting Gate) and a fixed submerged weir (Crump weir). Overall, the adjustable submerged weir consistently produces greater scour depth magnitudes, particularly at higher discharge levels. On average, the adjustable weir exhibits a deeper scour magnitude, with a maximum depth of -5.917 m at 3600 m³/s, compared to the fixed weir, which reaches a maximum depth of -5.128 m. This suggests that the hydraulic conditions generated by the adjustable weir are more conducive to intense local erosion, especially as flow increases. The mean scour depth for the adjustable submerged weir is visibly more negative across most discharge levels, indicating more substantial localised erosion. The fixed weir, in contrast, exhibits more moderate and fluctuating scour patterns, particularly in the low to mid-discharge range (up to 1800 m³/s), where its scour depth sometimes remains above -2,000 m. While the fixed submerged weir initially records shallower scour depth at lower flows for runs below 500 m3/s, it begins to narrow the scour depth gap with the adjustable weir as discharge increases beyond 2000 m^3/s. This shift could indicate a redistribution of flow energy across a broader area, as fixed weirs like the Crump type are known for diffusing hydraulic intensity more evenly downstream. These results underscore the influence of weir design on sediment transport and bed erosion. The adjustable weir design tends to concentrate flow energy more locally, leading to deeper and more focused scour holes, particularly under extreme flow conditions. Meanwhile, the fixed weir may encourage more distributed erosion patterns, potentially reducing the risk of structural undermining, but increasing the spread of sediment mobilisation zones.

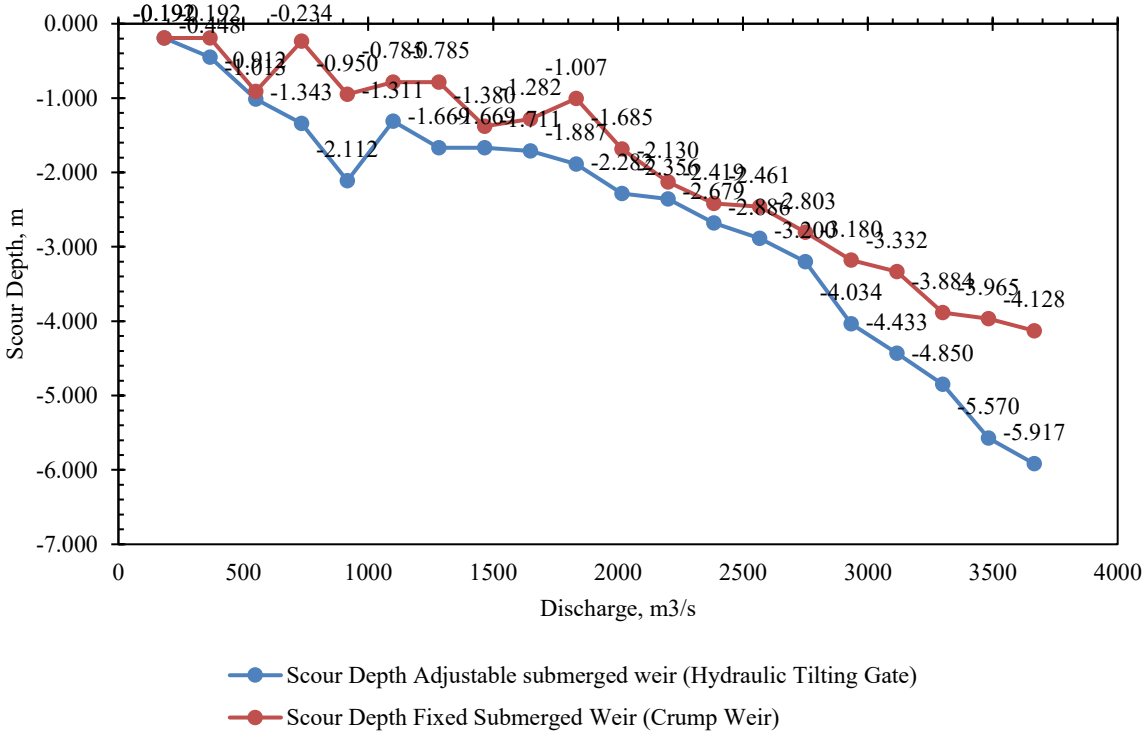


Figure 16. Scour Depth (Ds) versus Discharge (Q) for Adjustable Submerged Weir (Hydraulic Tilting Gate) and Fixed Submerged Weir (Crump Weir)

4. CONCLUSIONS

This numerical investigation using FLOW-3D demonstrates clear differences in sediment scour characteristics downstream of adjustable and fixed submerged weirs under varying flow

conditions. The results show that an adjustable submerged weir (hydraulic tilting gate) tends to produce more concentrated and deeper scouring. In contrast, a fixed submerged weir (Crump weir) generates broader, more distributed erosion patterns. In terms of flow velocity at scour depth, the adjustable weir recorded a higher average velocity of 4.848 m/s, compared to 4.300 m/s for the fixed weir across all simulated runs. The maximum velocity for the adjustable weir peaked at 5.152 m/s. In contrast, the fixed weir reached a higher peak velocity of 8.978 m/s under extreme discharge ($Q = 3666.00$ m^3/s), indicating its potential for intense flow dispersion and turbulence in high-flow events. Conversely, the minimum recorded velocity was 1.249 m/s for the adjustable weir and 0.629 m/s for the fixed weir, showing that the fixed weir had greater variability and potential for sediment deposition in low-energy zones.

Regarding scour depth, the adjustable weir consistently produced deeper scour holes, with a maximum depth of -5.917 m. compared to -5.128 m for the fixed weir. This pattern was most apparent at higher discharges, suggesting that the concentrated jet flow from the adjustable weir promotes localised erosion. The fixed weir, while showing shallower average scour depths in the early flow stages, begins to approach the scour levels of the adjustable weir as discharge increases beyond 2000 m^3/s, reflecting its broader hydraulic spread.

The findings highlight the importance of weir geometry and operational characteristics in influencing downstream erosion. Adjustable weirs provide hydraulic control but may intensify scour risks, while fixed weirs like the Crump type offer more predictable and distributed sediment transport behaviour, enhancing structural resilience under high-flow conditions. The data derived from this study can support future design improvements and sediment management strategies for submerged weirs in river engineering applications. However, this study is limited to numerical simulations without physical model validation for scouring results, and the scope is confined to clear water scour conditions without sediment supply from upstream. Future research should incorporate experimental validation, investigate sediment-laden flows, and assess the performance of other weir geometries under variable hydraulic and sediment conditions to improve the generalisability and applicability of the findings.

ACKNOWLEDGEMENTS

This study was funded through the IIUM Research Matching Grant Scheme (RMGS) under Project RMGS24-006-0037, titled “Numerical Modelling of Riverbed Scouring Upstream and Downstream of the Kasa Low Weir Structure, Pasir Mas, Across Sungai Kelantan.”

REFERENCES

[1] Lo, W. C., Su, H., & Shih, D. S. (2022). Integrating multiple downscaling simulations with continuous in-situ monitoring to assess riverbed scouring. Journal of Hydrology, 610, 127841. https://doi.org/10.1016/j.jhydrol.2022.127841

[2] Bodda, S. S., Gupta, A., Ju, B. S., & Jung, W. Y. (2020). Fragility of a weir structure due to scouring. Computational Engineering and Physical Modeling. https://doi.org/10.22115/cepm.2020.214539.1077

[3] Guan, D., Liu, J., Chiew, Y. M., & Zhou, Y. (2019). Scour evolution downstream of submerged weirs in clear water scour conditions. Water (Switzerland), 11(9). https://doi.org/10.3390/w11091746

[4] Guan, D., Melville, B. W., & Friedrich, H. (2015b). Scour at submerged weirs.

[5] Schleiss, A., De Cesare, G., Franca, M. J., & Pfister, M. (2014). River Flow 2014: Proceedings of the International Conference on Fluvial Hydraulics, Lausanne, Switzerland, 3–5 September 2014. CRC Press/Balkema.

[6] Zhang, W., Liu, X., & Gan, B. (2024). Experimental study on upstream water level rise of submerged rock weirs. Water (Switzerland), 16(15). https://doi.org/10.3390/w16152136

[7] Guan, D., Melville, B. W., & Friedrich, H. (2015a). Live-bed scour at submerged weirs. Journal of Hydraulic Engineering, 141(2). https://doi.org/10.1061/(ASCE)HY.1943-7900.0000954

[8] Guan, D., Melville, B., & Friedrich, H. (2016). Local scour at submerged weirs in sand-bed channels. Journal of Hydraulic Research, 54(2), 172–184. https://doi.org/10.1080/00221686.2015.1132275

[9] Ikinciogullari, E., Emiroglu, M. E., & Aydin, M. C. (2022). Comparison of Scour Properties of Classical and Trapezoidal Labyrinth Weirs. Arabian Journal for Science and Engineering, 47(4), 4023–4040. https://doi.org/10.1007/s13369-021-05832-z

[10] Seyedzadeh, A., Yasi, M., & Farhoudi, J. (2023). Hydraulic characteristics of a sluice gate over a broad-crested weir. Flow Measurement and Instrumentation, 93, 102432. https://doi.org/10.1016/j.flowmeasinst.2023.102432

[11] Seyedzadeh, A., Yasi, M., & Farhoudi, J. (2024). Investigating the flow through parallel sluice gates on a broad-crested weir. Flow Measurement and Instrumentation, 99, 102679. https://doi.org/10.1016/j.flowmeasinst.2024.102679

[12] Moradi, S., Noshahri, A. G., Barati, R., Esmaili, K., & Atashi, V. (2025). Experimental study of Crump weir for sediment transport and flow measurement in sediment-laden rivers. Scientific Reports, 15(1). https://doi.org/10.1038/s41598-025-00404-9

[13] Song, C. G., Park, S. W., & Shin, J. (2024). An experimental study of flow and turbulence properties near the rising sector gate mouth considering the gate opening with a PIV measuring system. Water (Switzerland), 16(20). https://doi.org/10.3390/w16203004

[14] Obaida, A. A. M., & Mohammed, A. Y. (2023). The effect of Crump weir's geometry changes on hydraulic flow characteristics: A review. International Journal of Advanced Natural Sciences and Engineering Researches, 8(8), 47–53. https://as-proceeding.com/index.php/ijanser

[15] Moradi, S., Noshahri, A. G., Barati, R., Esmaili, K., & Atashi, V. (2025). Experimental study of Crump weir for sediment transport and flow measurement in sediment-laden rivers. Scientific Reports, 15(1). https://doi.org/10.1038/s41598-025-00404-9

[16] AL-Naely, H., Al-Khafaj, Z., & Khassaf, S. (2018). Effect of opening holes on the hydraulic performance for Crump weir. International Journal of Engineering, 31(12). https://doi.org/10.5829/ije.2018.31.12c.05

[17] Chen, S. C., & Tfwala, S. S. (2018). Performance assessment of FLOW-3D and X flow in the numerical modelling of fish-bone type fishway hydraulics. In 7th IAHR International Symposium on Hydraulic Structures (ISHS 2018), 272–282. https://doi.org/10.15142/T3HH1J

[18] Jalal, H. K., & Hassan, W. H. (2020). Three-dimensional numerical simulation of local scour around a circular bridge pier using Flow-3D software. IOP Conference Series: Materials Science and Engineering, 745(1). https://doi.org/10.1088/1757-899X/745/1/012150

[19] Sim, Y. C., & Choi, S.-U. (2014). Three-dimensional scour at submarine pipelines under indefinite boundary conditions.

[20] Abd, R. M., & Rady, E.-H. (2011). 2D–3D modeling of flow over sharp-crested weirs. Journal of Applied Sciences Research, 7(12).

[21] Hamad, K., & Torres, C. (2025). Hydraulic analysis with CFD – FLOW-3D numerical modeling of local erosion in bridge piles. Journal of Building Technology, 7(1). https://doi.org/10.32629/jbt.v7i1.3410

[22] Priya Varshini, A. G., Anitha Kumari, K., Janani, D., & Soundariya, S. (2021). Comparative analysis of Machine learning and Deep learning algorithms for Software Effort Estimation. Journal of Physics: Conference Series, 1767(1). https://doi.org/10.1088/1742-6596/1767/1/012019

[23] Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science, 7, 1–24. https://doi.org/10.7717/PEERJ-CS.623

[24] Guo, X., Xu, L., Su, L., Deng, Y., & Yang, C. (2021). Comparing Flow Duration Curves and Discharge Hydrographs to Assess Eco-flows. Water Resources Management, 35(14), 4681–4693. https://doi.org/10.1007/s11269-021-02890-8

[25] Aqilah, F., Islam, M., Juretic, F., Guerrero, J., Wood, D., & Ani, F. N. (2018). Study of mesh quality improvement for CFD analysis of an airfoil. IIUM Engineering Journal, 19(2), 203–212. https://doi.org/10.31436/iiumej.v19i2.905

[26] Mojtahedi, A., Soori, N., & Mohammadian, M. (2020). Energy dissipation evaluation for stepped spillway using a fuzzy inference system. SN Applied Sciences, 2(8). https://doi.org/10.1007/s42452-020-03258-0

https://doi.org/10.31436/iiumej.v26i3.3433

DESIGNING AND IMPLEMENTING REAL-TIME DEEP LEARNING OBJECT DETECTION IN UNMANNED AERIAL VEHICLES

NURUL FAHMI ARIEF HAKIM [1], ALYANI ISMAIL [2], IWAN KUSTIAWAN [1*], DEASY ROSANTI [3], TAQIYUDDIN YAZID ZAIDAN [1]

[1]*Dept. of Electrical Engineering Education, Universitas Pendidikan Indonesia, Bandung, Indonesia*
[2]*Wireless and Photonic Networks Research Center, Universiti Putra Malaysia, Serdang, Malaysia*
[3]*Electrical Engineering Study Program, Politeknik TEDC, Bandung, Indonesia*

**Corresponding author: iwan_kustiawan@upi.edu*

(Received: 1 September 2024; Accepted: 29 June 2025; Published online: 9 September 2025)

ABSTRACT: The rapid development of surveillance technology is readily apparent, particularly in monitoring vast and remote locations that present difficulties for human accessibility. Within the realm of contemporary surveillance methods, the utilization of Unmanned Aerial Vehicles (UAVs) has garnered a considerable amount of interest. The swift advancements in science and technology have led to the progressive incorporation of Artificial Intelligence (AI) technologies, particularly in tasks such as monitoring and reconnaissance. The results of this research contribute to the creation of a prototype for a UAV capable of conducting autonomous monitoring missions and integrating artificial intelligence technologies for real-time video processing. This study utilized an experimental methodology, and a Raspberry Pi was utilized for artificial intelligence processes integrated with the aircraft controller. During the decisive experiment, the unmanned aerial vehicle UAV could effectively travel to the designated area, reaching an accuracy of 78.6% in its AI processing.

ABSTRAK: Perkembangan pesat teknologi pemantauan kini semakin ketara, terutamanya dalam pemantauan kawasan yang luas dan terpencil sukar diakses manusia. Dalam konteks kaedah pemantauan kontemporari, penggunaan Kenderaan Udara Tanpa Pemandu (UAV) telah menarik minat ramai. Kemajuan pesat sains dan teknologi telah menghasilkan gabungan teknologi Kecerdasan Buatan (AI) secara progresif, terutama dalam bidang pemantauan dan peninjauan. Penyelidikan ini memberi sumbangan kepada penciptaan prototaip UAV yang mampu menjalankan misi pemantauan sendiri, disamping gabungan teknologi kecerdasan buatan bagi memproses video masa nyata. Kajian ini menggunakan metodologi eksperimen dan Raspberry Pi yang disepadukan dengan pengawal pesawat bagi proses kecerdasan buatan. Semasa eksperimen penentuan, UAV berjaya bergerak dengan berkesan ke kawasan yang ditetapkan, mencapai ketepatan 78.6% dalam pemprosesan AI.

KEYWORDS: *Autonomous vehicle, deep learning, object detection, surveillance, UAV.*

1. INTRODUCTION

Unmanned Aerial Vehicles (UAVs) have experienced rapid development and adoption across various sectors, including agriculture, search and rescue, surveillance, transmission line maintenance, infrastructure inspection, environmental mapping, and more [1]. Their flexibility, mobility, and cost-effectiveness have made UAVs indispensable for both civilian and military operations, particularly in ground monitoring and surveillance tasks [2]. Among their key

https://doi.org/10.31436/iiumej.v26i3.3433

advantages are real-time video transmission, high-resolution imaging, and the ability to operate in complex or high-risk environments [3]. However, most existing UAV-based surveillance systems are limited by their reliance on manual operation and post-processing, which constrains their ability to deliver timely, autonomous threat detection and response.

As urban environments become increasingly complex and densely populated, there is an urgent need for intelligent, adaptive, and real-time aerial surveillance systems to enhance public safety and situational awareness [4]. Traditional systems often struggle to respond effectively to the dynamic and unpredictable nature of urban threats. Static cameras and manually controlled drones are insufficient for ensuring fast and reliable surveillance, especially in densely populated or hard-to-reach areas [5]. Moreover, with rising incidents of traffic congestion, public safety threats, and emergency events, surveillance platforms must be able to detect and respond to anomalies autonomously. The integration of artificial intelligence into UAVs offers a promising solution, shifting surveillance from passive observation to proactive and autonomous monitoring [6].

Prior studies have explored AI integration in UAVs, such as video surveillance analysis using the Aids to Navigation (ATON) dataset and vehicle speed estimation through deep neural networks [7], [8]. The videos utilized are sourced from the ATON dataset, which comprises numerous surveillance camera and UAV footage that is subsequently processed by AI algorithms. Nevertheless, the level of user practicality is reduced when the AI system is not real-time and is applied while the UAV is in operation. This research addresses that need by developing an AI-enhanced UAV system capable of both manual and autonomous flight, designed explicitly for real-time surveillance and object detection. By equipping the UAV with a microcomputer and onboard camera, the system integrates AI algorithms to automatically detect and recognize objects in real time, particularly near residential zones and roadways. Unlike conventional systems that require manual monitoring or offline video analysis [9], the proposed system functions autonomously and processes visual data onboard, reducing latency and operator dependency.

The uniqueness of this article is its comprehensive integration of autonomous navigation, integrated AI-powered visual processing, and real-time object recognition within a single UAV platform. This method not only enhances the current capabilities of UAVs but also addresses the urgent need for intelligent, self-sufficient aerial monitoring solutions. The contribution of this article is to introduce a practical, scalable, and cost-effective UAV surveillance framework that utilizes lightweight AI models and accessible hardware components, making it feasible for broader adoption in various applications. The software is also open-source, allowing everyone to contribute to its development. Moreover, it enhances the state-of-the-art UAV-based surveillance by enabling proactive monitoring and immediate object detection without ground-based computation or human intervention. By bridging the gap between UAV mobility and artificial intelligence, this study contributes to advancing intelligent sensing technologies, with significant implications for innovative city development, disaster response systems, and real-time traffic and security monitoring.

2. METHOD

The methodology for designing and implementing UAVs follows the waterfall approach, as illustrated in Fig. 1. This approach adheres to a pre-established sequence and linear progression, where each phase builds upon the previous one. The UAV's design process begins with gathering requirements to ensure the UAV conforms to specified criteria. In the second phase, the system design occurs, encompassing the architecture of the UAVs, including

https://doi.org/10.31436/iiumej.v26i3.3433

hardware, software, and integration with a deep learning object detection system. During the implementation phase, the hardware and software components of the UAV are constructed and developed. Using 3D design software, the UAV's frame is built, taking into account the flight controller, GPS, sensors, cameras, motors, and propellers. Concurrently, the deep learning object detection system is integrated with the onboard hardware for real-time processing. Subsequent phases focus on constructing a three-dimensional model of the UAVs and designing hardware and software systems, ensuring the integration of all essential components. The deep learning procedure captures object images to generate training data for machine learning models. A functional prototype is then constructed by integrating UAV subsystems and object detection systems, enabling the UAVs to navigate autonomously during testing. Once the system integration is complete, the UAV prototype will detect objects during monitoring missions near residential areas or roadways. The system's performance is thoroughly evaluated, and predetermined criteria are applied to ensure its capability for monitoring aerial objects and the proper functioning of the deep learning program.

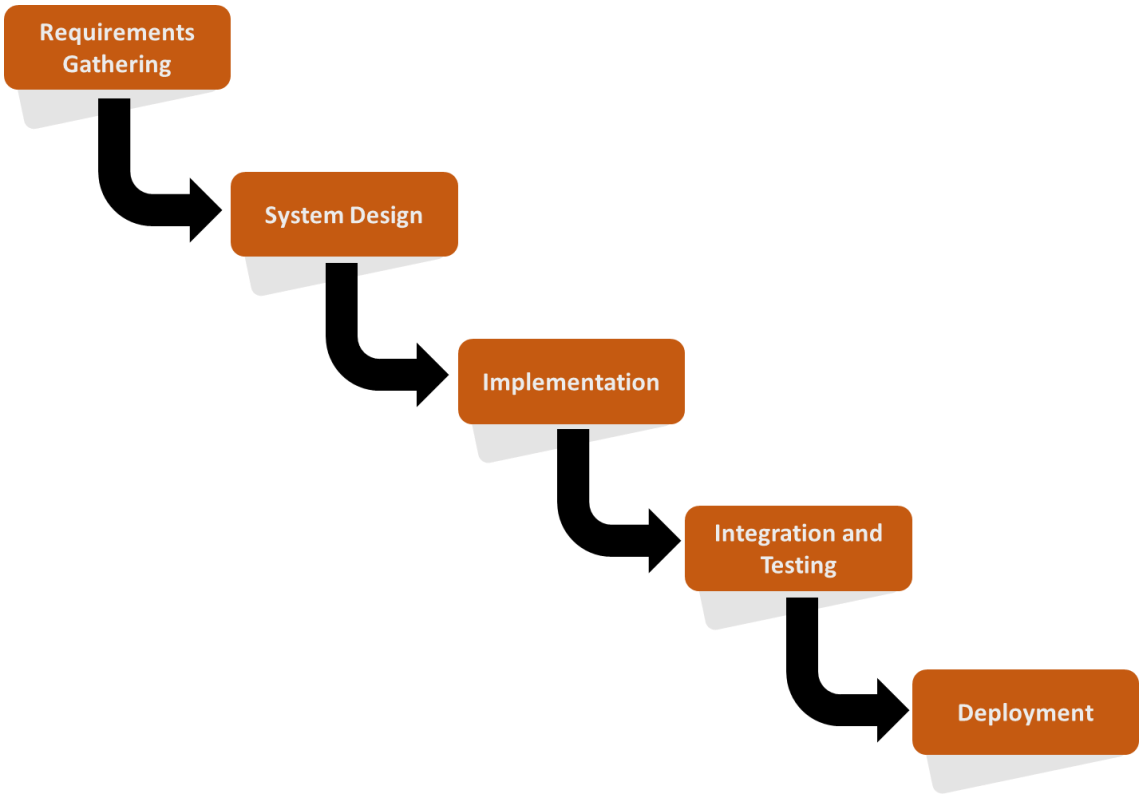


Figure 1. UAVs design flowchart

2.1. UAVs design

The proposed UAV design features the F330 frame, which has a diagonal length of 330 mm and a height of 70 mm. To accommodate the addition of an AI system, extra design parts will be implemented on the legs, resulting in an increased height of the UAVs [10]. The body of the UAV design is illustrated in Fig. 2. In the hardware design, this UAV's control system utilizes the Pixhawk Flight Controller. The Pixhawk controller features a built-in inertial measurement unit (IMU) module with four sensors: a magnetometer, a gyroscope, a barometer, and an accelerometer. In addition to the Pixhawk controller, the UAV's frame is equipped with GPS, a radio control receiver, a battery, a camera, and other components. The types of UAV system components and their respective positions are detailed in Table 1.

https://doi.org/10.31436/iiumej.v26i3.3433

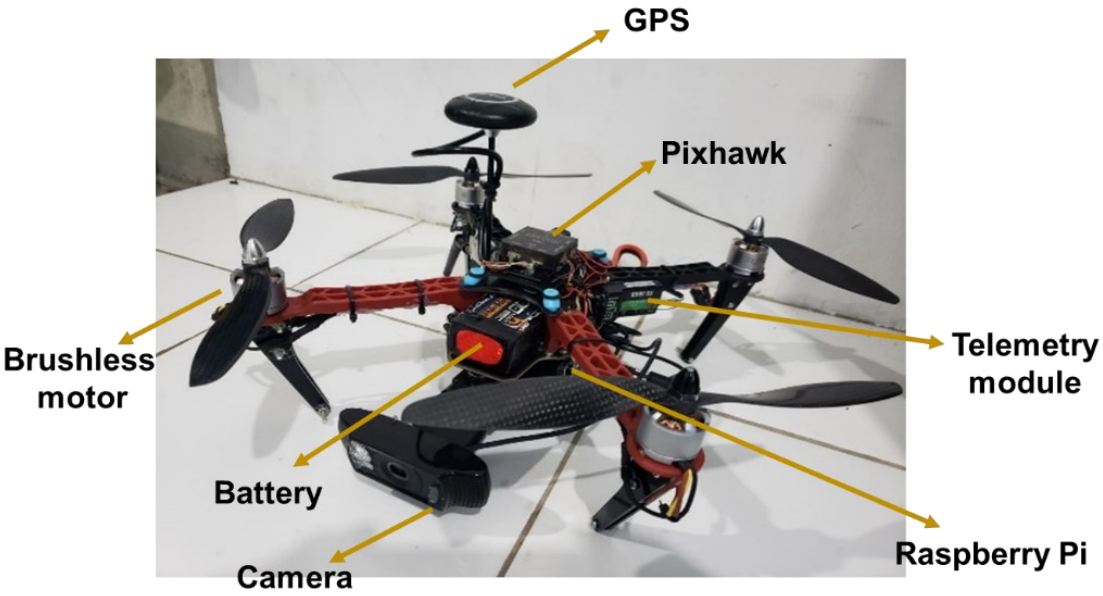


Figure 2. The design of the part increases the height of the landing

Table 1. UAV quadcopter components

Components	Description	
	Type	Position
Frame	F330 quadcopter frame	Main
Flight Controller	Mini Pixhawk PX4	Top layer (center)
GPS	Neo GPS M8N Module	Top layer (center with holder)
Radio Control Receiver	FLSky I6S	Bottom layer (rear left)
Battery	Dinodigy 5000mAh 3S 11,1 Volt 65C	Middle (between the top and bottom layers)
Electronic Speed Controller	4 ESCs BLHeli 25A	Each arm (bottom)
Telemetry for flight data acquisition	2.433 MHZ Radio Telemetry 500mA	Bottom layer (rear right)
Motor	4 Brushless motor 2212 920KV	Each arm
Propeller	4 Propeller Carbon 9047	Each arm
Single Board Computer	Raspberry Pi 4 Model B	Bottom layer (center)
Camera	Logitech C920 Webcam	Bottom layer (front)
Miscellaneous	RC PPM encoder	Ground Control Station
	Power module	Bottom layer
	Buzzer	Top layer
	Pixhawk damping plate	Top layer
	GPS folding antenna	Top layer
	Camera shock absorber	Bottom layer

The UAV hardware block diagram is shown in Fig. 3. The UAV's hardware design centers around two control systems: the Pixhawk, which handles flight processing, and the Raspberry Pi, which manages AI processing. The Pixhawk control system is connected to the GPS, which provides navigation for the UAV. Integrated GPS/INS (Inertial Navigation System) navigation is the standard approach in integrated navigation research. Additionally, the Pixhawk is connected to four ESCs (Electronic Speed Controllers) that regulate four brushless DC motors. Unlike traditional DC motors, brushless DC motors do not have brushes, offering benefits such as high efficiency, speed, torque, and easy maintenance [11]. The Pixhawk also communicates

https://doi.org/10.31436/iiumej.v26i3.3433

with the Ground Control Station (GCS) via telemetry. The GCS is the central hub for UAV ground control and is the only means ground personnel can communicate with the UAV. It functions as a control system, monitoring the UAV and creating flight missions to enable autonomous operation. Furthermore, the Pixhawk is connected to a remote receiver module, which receives input signals from the radio control during manual flights.

Mission Planner is a ground control software frequently employed to establish communication and configuration with Pixhawk, an open-source flight controller for unmanned aerial vehicles (UAVs). Mission Planner allows the UAV to operate autonomously when embedded and configured correctly, executing predetermined flight missions without real-time manual control. Mission Planner is embedded with Pixhawk and is utilized as an autonomous UAV system. Based on real-time sensor inputs (GPS, accelerometer, gyroscope), the flight controller controls the quadcopter in autonomous mode to follow a pre-programmed flight path or execute tasks such as waypoint navigation, takeoff, landing, or hovering. This configuration enables the UAV to function autonomously, even in environments where human control may be challenging or inefficient. An AI algorithm is created to process the video or sensor data of the UAV using the Python programming language once reliable data has been obtained. The AI model can be customized to perform various duties, including object detection, tracking, and classification, contingent upon the mission's objectives. This flight data undergoes processing for filtering and synchronization [12]. The development of the AI model is conducted using Google Colaboratory (Colab), a free cloud-based environment that enables researchers to write and execute Python code in the browser. This is to train and validate the model [13].

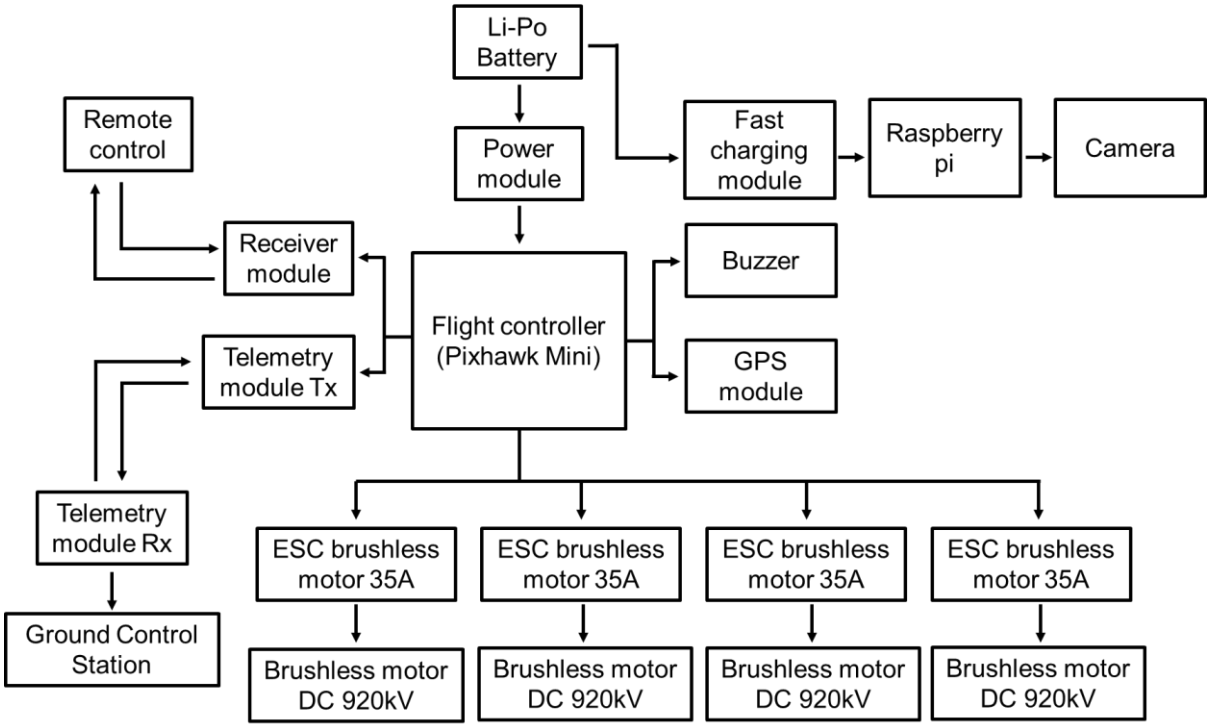


Figure 3. Block diagram of the hardware UAV

2.2. Object detection system

For arithmetic calculations in machine learning, Google has developed an open-source tool called TensorFlow. TensorFlow offers significant advantages in terms of high accessibility and flexibility, with ample resources available on GitHub and online forums to enhance its

https://doi.org/10.31436/iiumej.v26i3.3433

efficiency. Currently, Google provides several pre-trained models on the official TensorFlow website, and MobileNets is one such model trained within TensorFlow. MobileNets is a convolutional neural network (CNN) architecture designed to address the need for efficient computing resources [14]. TensorFlow facilitates transfer learning, a novel machine learning approach that leverages existing knowledge, learning from one environment to solve new, different, but related problems. This is particularly valuable when there is insufficient training data or a need to train a custom model without starting from scratch [15]. The object detection system developed in this research utilizes the TensorFlow API. In this study, researchers employed transfer learning methods to train a custom dataset using the MobilenetV2 model, creating a customized machine-learning model. The objects targeted for training comprise four classes: cars, motorcycles, trees, and trucks.

2.3. Data Preparation

The utilized data comprises a collection of images captured from both cell phone and UAV cameras. Examples of image data are presented in Fig. 4. These images are extracted into frames in *JPG format. They will serve as samples in image processing for detecting roads, cars, motorcycles, trees, and trucks. The total number of collected images amounts to 1,400. Each image will undergo labeling for object detection, as illustrated in Fig. 5. Data labeling has been a focal point in machine learning research. Subsequently, after labeling, the data will be partitioned into 80% training data, 10% validation data, and 10% test data.

Figure 4. Sample picture on the main road

https://doi.org/10.31436/iiumej.v26i3.3433

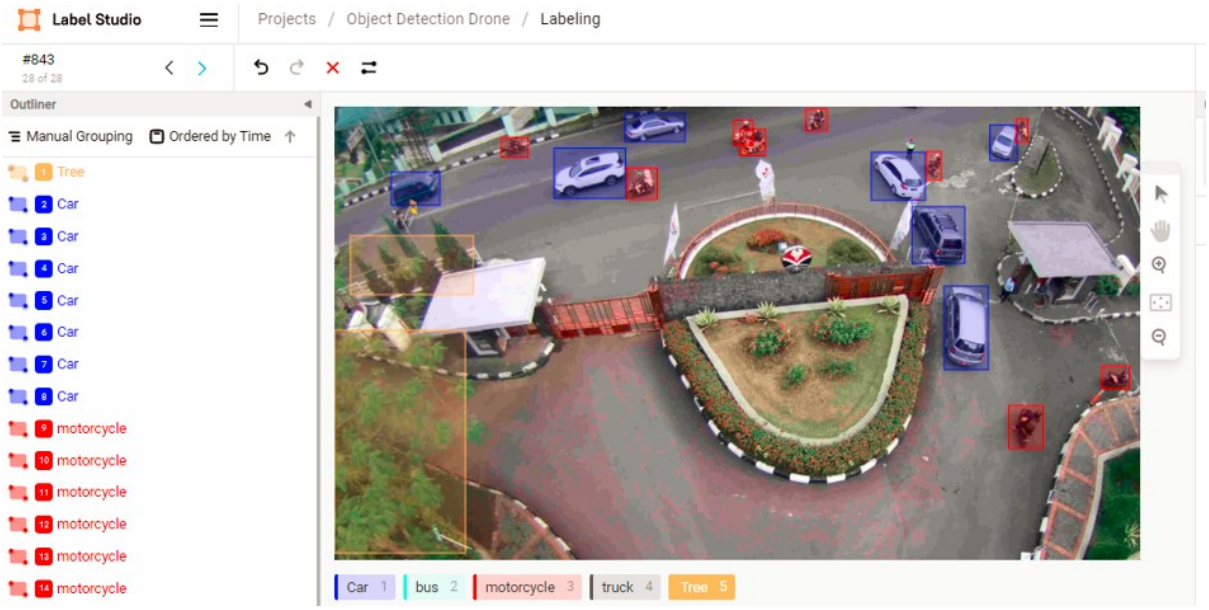


Figure 5. Data labeling using Label Studio

2.4. Deep learning

Deep learning is a machine learning algorithm component used to manage large datasets. Deep learning is swiftly expanding and gaining momentum to resolve digital image and video issues. In recent years, deep learning has been implemented to facilitate the classification of images and the detection of objects. Equations (1) and (2). The object detection system on a UAV will be developed using a deep learning algorithm. The TensorFlow API will be employed by the object detection system developed in this study to generate a deep learning model. The transfer learning method trains a custom dataset from the MobilenetV2 model into a new custom deep learning model.

The construction of a deep learning model is divided into three phases, as illustrated in Figure 6. The following stages are as follows: data preparation, training, and testing [16]. The data preparation stage is the initial phase of the deep learning training process. These stages encompass data acquisition, labeling, LabelMap configuration, and converting the data label results into a pipeline for processing in the subsequent stage. Additionally, the training process stage will require significant time due to the computer's learning process. This stage comprises the convolution, pooling, and activation ReLU stages, which connect all process layers and generate a classification model. The loss value will be determined after the machine's learning or training process. If the Loss value is suitable, the process will proceed to the subsequent stage. If the results are unsatisfactory, the training process stage will be reverted to the training stage by adjusting several parameters to achieve the desired outcome. During the assessment process, the model undergoes a trial by inputting the resulting images and videos. Object detection results will be visible after the testing procedure, allowing conclusions to be drawn.

$$Accuracy = \frac{Images\ Validation\ (True\ Detected\ Object)}{Number\ of\ Total\ Images\ Validation}) \quad (1)$$

$$Loss = \frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{k}(Images\ Train - Images\ Validation)^2 \quad (2)$$

https://doi.org/10.31436/iiumej.v26i3.3433

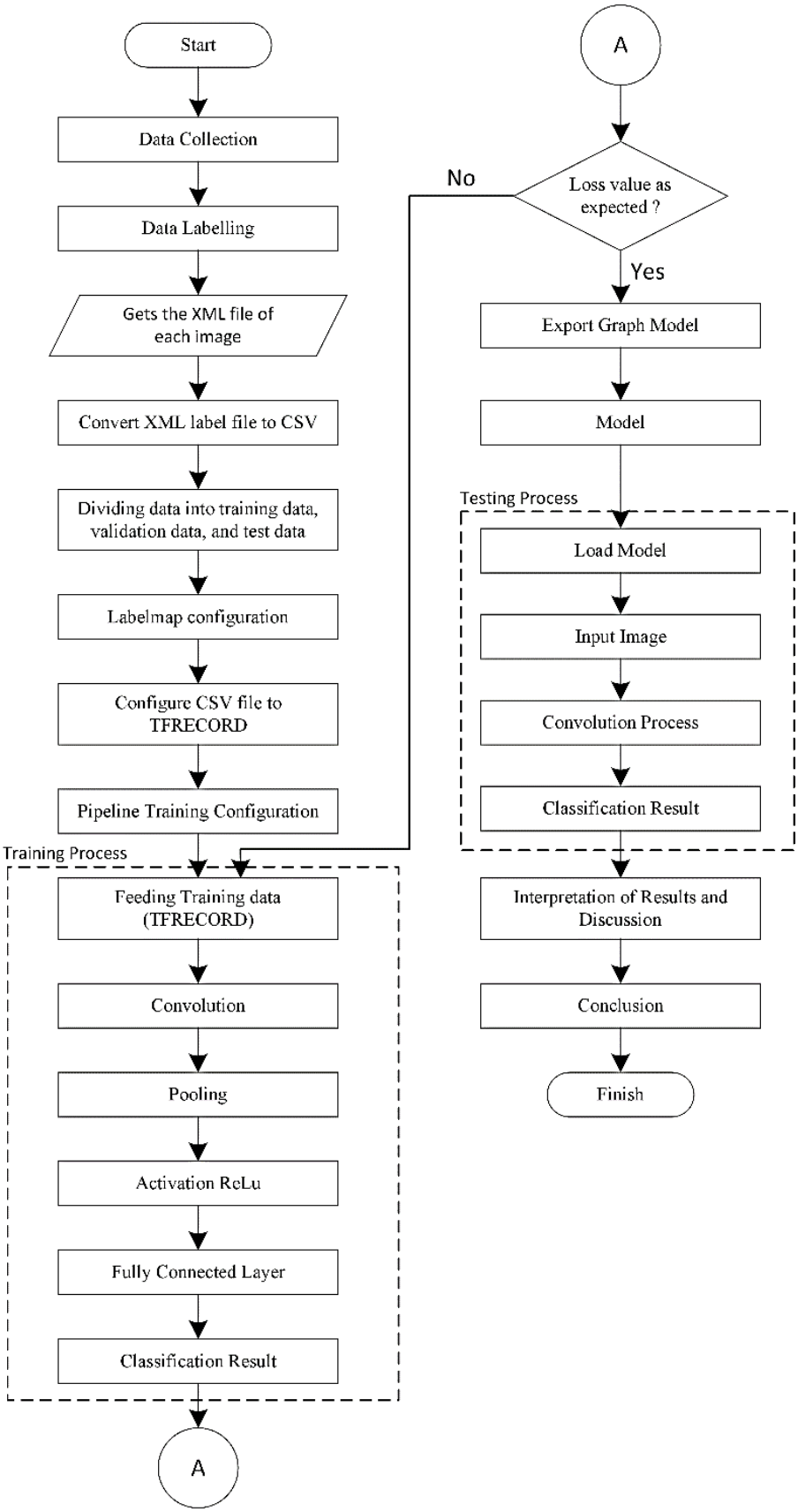


Figure 6. Deep learning flowchart

https://doi.org/10.31436/iiumej.v26i3.3433

3. RESULT AND DISCUSSION

The data taken in this study are in the form of traffic videos on Setiabudhi Street, Bandung, traffic in the Cimahi residential area, and flyovers. After the data is collected, labeling and training are carried out to get the desired machine-learning model. The result of the labeling is in the form of an XML file, which contains the coordinates and bounding box characteristics of each object from each image, as shown in Fig. 7. The number of images taken is 1400. Each image will undergo labeling and subsequent division into 80% (1120 images) for training data, 10% (140 images) for validation data, and 10% (140 images) for test data. The actual images used to train the model are called training data. The neural network is fed images from the "train" set at each training phase. The network forecasts the class and location of objects in the image. The training algorithm modifies the network weights through back-propagation and calculates the loss, which indicates the degree of "loss" in the predictions. The training algorithm utilizes validation data, which comprises images from a "validation" set, to monitor training progress and adjust hyperparameters, including learning rate. Validation images are employed intermittently during training, typically once every specified number of training steps, in contrast to the images in the "train" set. The test data comprises images that the neural network has never encountered during training. These images are intended for human inspection to conduct the final testing of models and assess their accuracy.

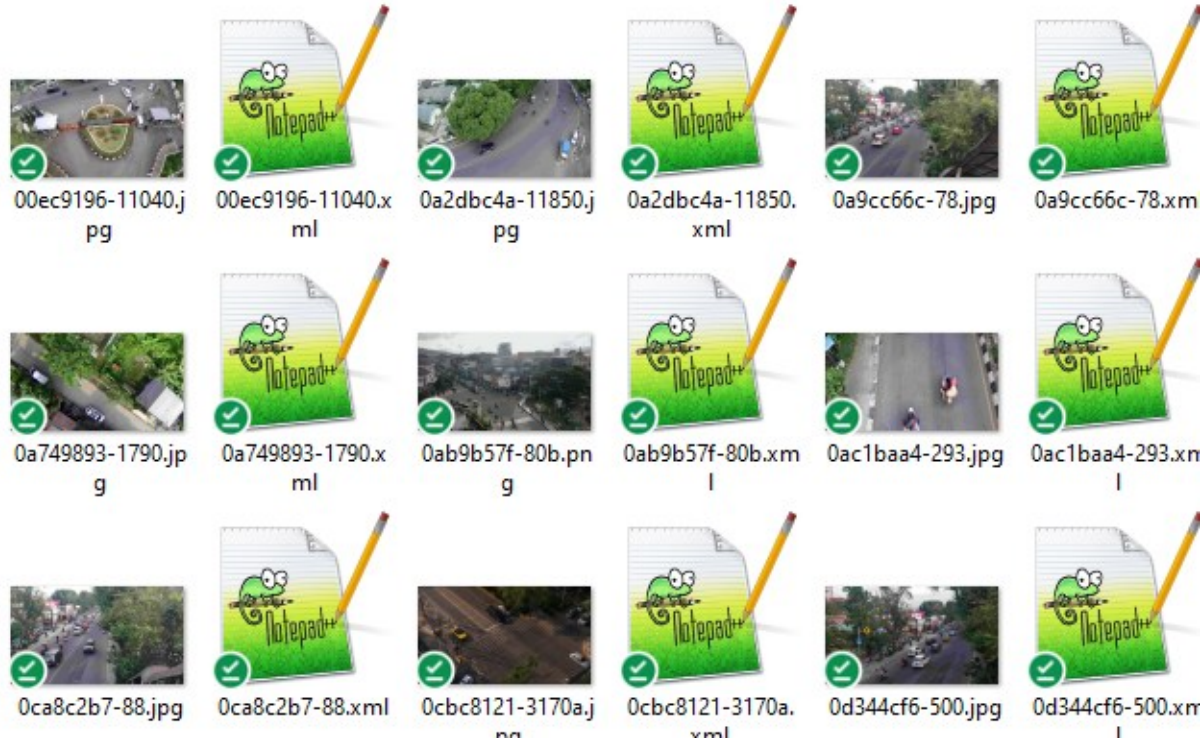


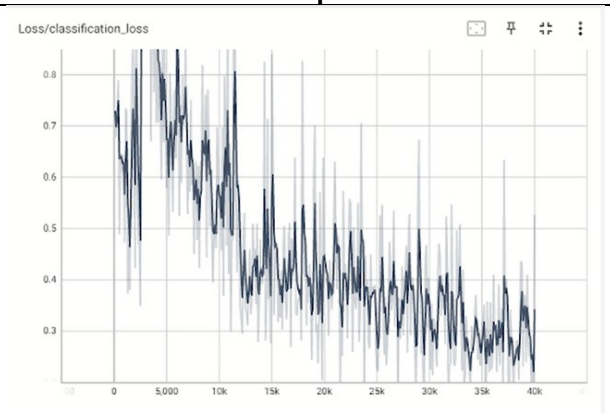
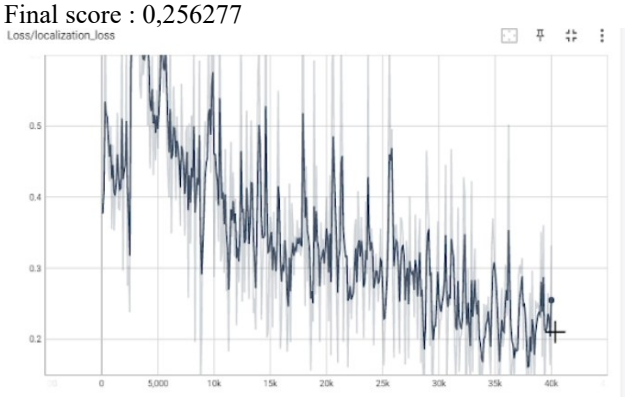
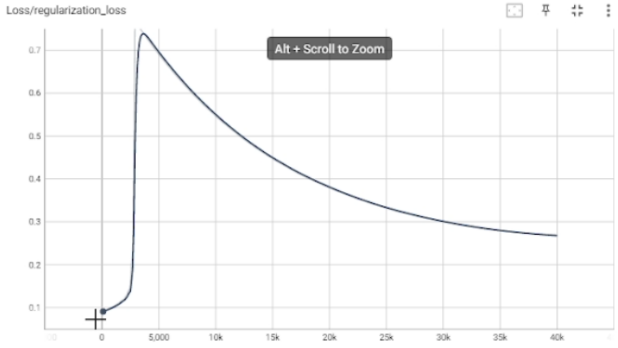
Figure 7. Labelling Result JPG and XML files for each image

In this AI algorithm, objects have been identified, and a label map needs to be created. The label map lists predefined classes: car, motorcycle, tree, and truck. Additionally, each data point in the training and validation sets will be compiled into a .csv file containing all images and their annotations, creating a TFRecord. The TFRecord file is designated for the training and validation datasets, alongside the label map.pbtxt file, encompassing label maps in various formats. Once the data is prepared for training, the training configuration for the SSD-MobileNetV2 model is established, which will be used for training.

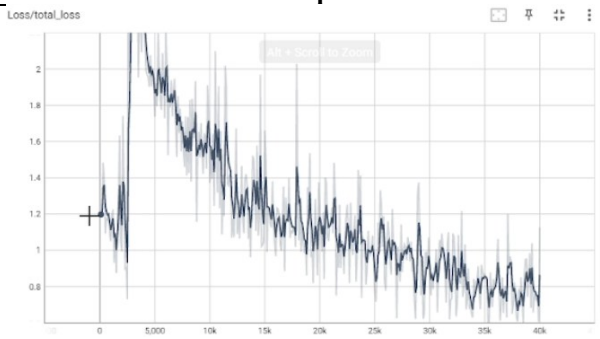
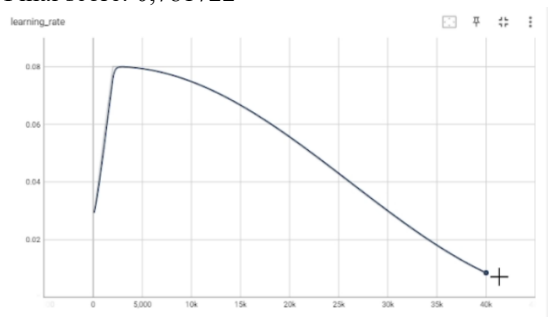
https://doi.org/10.31436/iiumej.v26i3.3433

When the configuration file is prepared, the subsequent stage is to specify the 'batch_size' and 'num_step'. The model will be trained using a total of 'num_step' steps. It is advised to commence with 40,000 steps. The duration of the training increases as the number of strides increases. The quantity of images employed in each training phase is denoted by 'batch_size'. The model can be trained in fewer stages with a larger batch size; however, the size is constrained by the available GPU memory for training. Additionally, the object detection training is prepared for execution. The training process will take 2-6 hours, depending on the GPU used for training, model, 'batch_size', and 'num_step', and it utilizes scripts from the TF Object Detection API. Table 2 displays the training results after the model has been effectively trained. The training results on the graph indicate that the loss is reduced and the information is more detailed as the number of steps increases.

Table 2. Training Result

Loss	Graphics	Description
Classification Loss	 Final score : 0,256277	Classification Loss shows whether the bounding box class matches the predicted class.
Localization Loss	 Final score : 0,257642	Localization Loss calculates the difference between bounding box predictions and labels.
Regularization Loss	 Final score: 0,267803	The network regularization function generates Regularization Loss and helps point optimization algorithms in the right direction.

https://doi.org/10.31436/iiumej.v26i3.3433

Loss	Graphics	Description
Total Loss	 Final score: 0,781722	Total Loss is the sum of the three previous Losses
Learning Rate	 Final score: 0,008266	The Learning Rate controls how much the weights are updated according to the estimated error

The results of object detection are presented in Figs. 8 and 9. Objects in images and videos are successfully identified with bounding boxes, image labels, and the percentage of image similarity. Based on the detection results, the AI algorithm does not detect several objects. Factors that affect the error include the object being tiny and other objects blocking the view of the object. Furthermore, the training model will be tested in real time when the UAV operates on a monitoring mission. carrying out the mission, it is necessary to make waypoints as shown in Fig.10. The waypoint navigation system is crafted to enable the drone to identify its position and gauge the distance it has covered, enhancing precision in reaching the designated destination along the predetermined route set by the operator.

Figure 8. Object detection from a picture: UAV detected 7 cars, and 1 car was not detected

https://doi.org/10.31436/iiumej.v26i3.3433

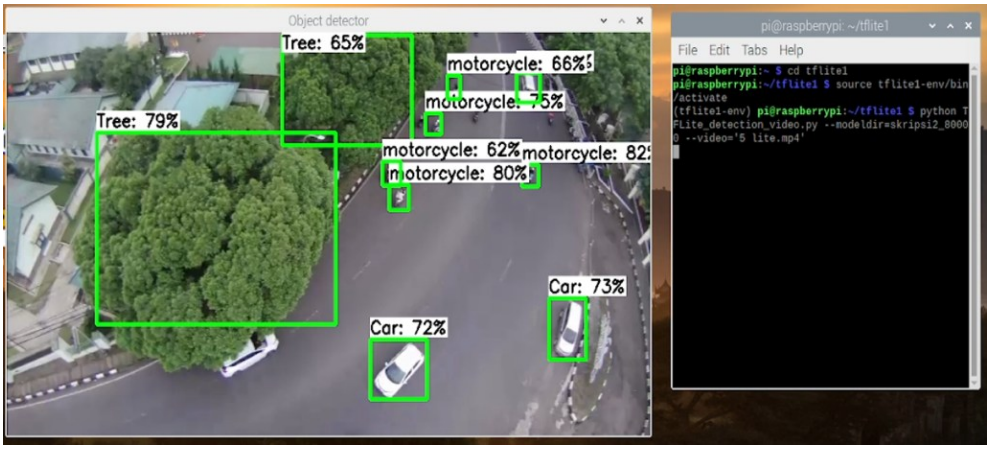


Figure 9. Object detection from video: UAV detected 2 cars, 6 motorcycles, 2 trees, and 1 car was not detected because it was covered by trees

Figure 10. Autonomous flying missions

The results of the autonomous UAV's real-time surveillance mission, specifically related to object detection, are illustrated in Figs. 11 and 12. The UAV descended safely to the last waypoint, exhibiting real-time object detection results aligned with those detected in image and video analysis. Successful object identification during the autonomous UAV mission is indicated by bounding rectangles containing labels specifying the class and the similarity percentage. It is worth mentioning that the Raspberry Pi, which is utilized to process AI for real-time object detection, exhibits a frames-per-second (FPS) value falling below 2. Standard video, on the other hand, generally attains 20-30 FPS. The inconsistency emerges because the Raspberry Pi, which powers the AI algorithm that processes the video, lacks a Graphics Processing Unit (GPU), an essential element in image processing.

Accuracy is assessed throughout the testing process by comparing real objects and those detected by the program. Achieving an accuracy of 78.6%, the object detection algorithm accurately classified 96 out of a possible 122 objects in an image comprising automobiles, motorcycles, trees, and vehicles during a test. The total number of objects was 122. Object detection errors can be caused by a multitude of factors, including the color of the vehicles.

https://doi.org/10.31436/iiumej.v26i3.3433

The absence of color diversity among the vehicles utilized as objects in this study has led to the need for more varied datasets. Additionally, vehicles positioned so far away that they appear as specks were labeled during image labeling, resulting in the model being trained to identify objects whose labels did not correspond to the actual ones.

Figure 11. The UAV detected 2 cars and 1 tree

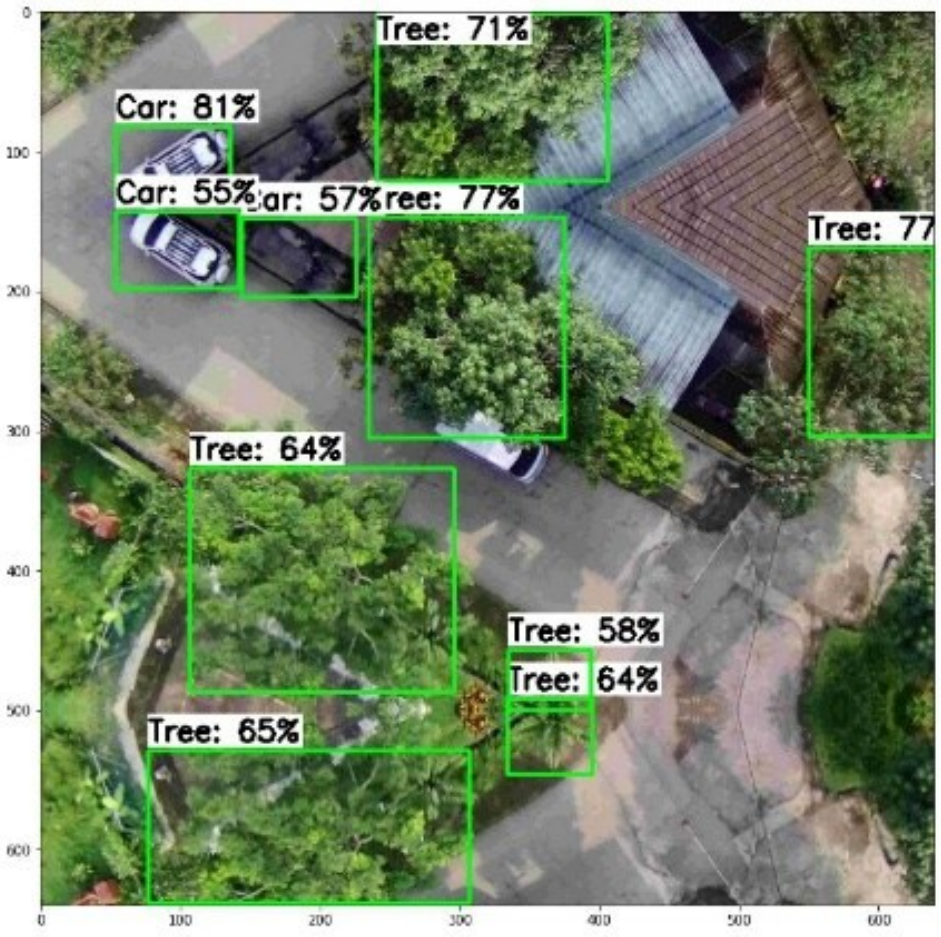


Figure 12. The UAV detected 3 cars and 7 trees

https://doi.org/10.31436/iiumej.v26i3.3433

Integrating Amazon Web Services (AWS) for real-time UAV monitoring demonstrates significant potential for advancing image sensing technology in today's trends and challenges. In the current landscape, UAVs with advanced imaging sensors are widely used for applications such as precision agriculture, environmental monitoring, disaster response, and infrastructure inspection. Fig.13 illustrates the website's monitoring capabilities. The website displays coordinate data that indicates the UAV's altitude is 30.64 meters, which is consistent with the mission. The longitude and latitude of the UAV are also indicated. The UAV has a ground speed of 2.247 and an air speed of 3.579. The currently displayed mode is also suitable, specifically AUTO Mode.

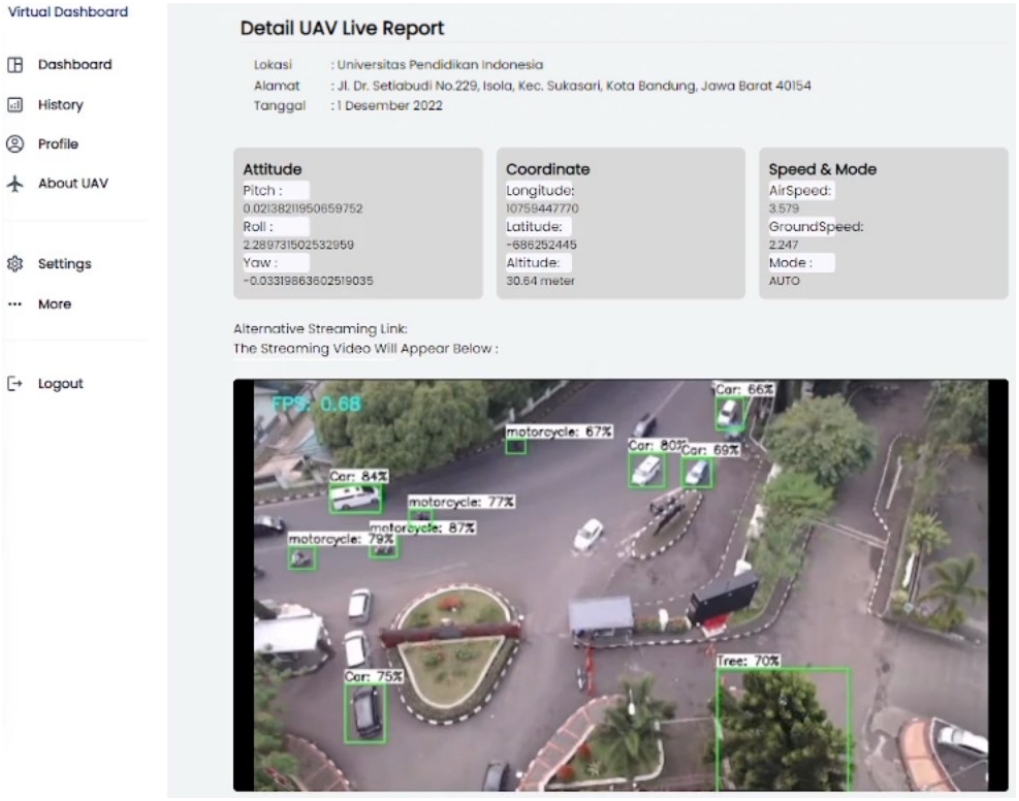


Figure 13. UAVs' real-time monitoring website

4. CONCLUSION

Real-time object detection utilizing deep learning and custom datasets has been successfully developed. The UAV successfully navigated to the last waypoint while simultaneously conducting AI processes. The trained model has proven effective in predicting objects in images and videos captured by the UAV and can operate in real-time during UAV operations. The actual detection rate reaches only 78.6% when the UAV is used for object detection. Detection errors stem from limited datasets and imperfect labels. A more substantial GPU is required to enhance AI processing performance in real-time (increase FPS). For instance, one could integrate a USB Coral for an external GPU with a Raspberry Pi, or opt for a more powerful single-board computer, such as the Nvidia Jetson.

REFERENCES

[1] S. Byun, I.-K. Shin, J. Moon, J. Kang, and S.-I. Choi, "Road Traffic Monitoring from UAV Images Using Deep Learning Networks," Remote Sens (Basel), vol. 13, no. 20, p. 4027, Oct. 2021, doi: 10.3390/rs13204027.

[2] S. A. H. Mohsan, N. Q. H. Othman, Y. Li, M. H. Alsharif, and M. A. Khan, "Unmanned aerial vehicles (UAVs): practical aspects, applications, open challenges, security issues, and future trends," Intell Serv Robot, vol. 16, no. 1, pp. 109–137, Mar. 2023, doi: 10.1007/s11370-022-00452-4.

[3] A. Fascista, "Toward Integrated Large-Scale Environmental Monitoring Using WSN/UAV/Crowdsensing: A Review of Applications, Signal Processing, and Future Perspectives," Sensors, vol. 22, no. 5, p. 1824, Feb. 2022, doi: 10.3390/s22051824.

[4] J. C. N. Bittencourt, D. G. Costa, P. Portugal, and F. Vasques, "A Survey on Adaptive Smart Urban Systems," IEEE Access, vol. 12, pp. 102826–102850, 2024, doi: 10.1109/ACCESS.2024.3433381.

[5] I. Mademlis, C. Symeonidis, A. Tefas, and I. Pitas, "Vision-based drone control for autonomous UAV cinematography," Multimed Tools Appl, vol. 83, no. 8, pp. 25055–25083, Aug. 2023, doi: 10.1007/s11042-023-15336-7.

[6] N. Jain, A. Gambhir, and M. Pandey, "Unmanned Aerial Networks—UAVs and AI," 2024, pp. 321–351. doi: 10.1007/978-981-97-6790-8_12.

[7] M. T. Nguyen, L. H. Truong, and T. T. H. Le, "Video Surveillance Processing Algorithms utilizing Artificial Intelligent (AI) for Unmanned Autonomous Vehicles (UAVs)," MethodsX, vol. 8, p. 101472, 2021, doi: 10.1016/j.mex.2021.101472.

[8] A. Gomaa, T. Minematsu, M. M. Abdelwahab, M. Abo-Zahhad, and R. Taniguchi, "Faster CNN-based vehicle detection and counting strategy for fixed camera scenes," Multimed Tools Appl, vol. 81, no. 18, pp. 25443–25471, Jul. 2022, doi: 10.1007/s11042-022-12370-9.

[9] C. Chalmers, P. Fergus, C. A. Curbelo Montanez, S. N. Longmore, and S. A. Wich, "Video analysis for the detection of animals using convolutional neural networks and consumer-grade drones," J Unmanned Veh Syst, vol. 9, no. 2, pp. 112–127, Jun. 2021, doi: 10.1139/juvs-2020-0018.

[10] T. Y. Zaidan, N. F. A. Hakim, I. Kustiawan, A. M. Ridwan, S. A. T. Al Azhima, and I. Fahmi, "Video Live Streaming System on Unmanned Aerial Vehicles (UAVs) Using Autonomous Waypoint Navigation to Support Traffic Monitoring System," in 2023 9th International Conference on Wireless and Telematics (ICWT), IEEE, Jul. 2023, pp. 1–5. doi: 10.1109/ICWT58823.2023.10335464.

[11] S. Kumar Vishwakarma and M. Rahman, "Speed Control of PV Array-Based Z-Source Inverter Fed Brushless DC Motor Using Dynamic Duty Cycle Control," in 2022 2nd International Conference on Emerging Frontiers in Electrical and Electronic Technologies (ICEFEET), IEEE, Jun. 2022, pp. 1–6. doi: 10.1109/ICEFEET51821.2022.9847935.

[12] D. F. Carlson and S. Rysgaard, "Adapting open-source drone autopilots for real-time iceberg observations," MethodsX, vol. 5, pp. 1059–1072, 2018, doi: 10.1016/j.mex.2018.09.003.

[13] Dr. S. R. Sukhdeve and S. S. Sukhdeve, "Google Colaboratory," in Google Cloud Platform for Data Science, Berkeley, CA: Apress, 2023, pp. 11–34. doi: 10.1007/978-1-4842-9688-2_2.

[14] M. Dou, Z. Hong, and M. Shi, "An Improved Efficient Convolutional Neural Network for Weed Seedlings Detection," in 2021 International Conference on Culture-oriented Science & Technology (ICCST), IEEE, Nov. 2021, pp. 285–289. doi: 10.1109/ICCST53801.2021.00067.

[15] Y. Roh, G. Heo, and S. E. Whang, "A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective," IEEE Trans Knowl Data Eng, vol. 33, no. 4, pp. 1328–1347, Apr. 2021, doi: 10.1109/TKDE.2019.2946162.

[16] S. K. Baduge et al., "Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications," Autom Constr, vol. 141, p. 104440, Sep. 2022, doi: 10.1016/j.autcon.2022.104440.

https://doi.org/10.31436/iiumej.v26i3.3471

AN INTEGRATION OF LENET WITH REGULARIZATION TECHNIQUES FOR ELECTRONIC NOSE IN AIR CONTAMINANT CLASSIFICATION

MUHAMMAD ARIF SYAHMI MD DZAHIR, KIM SENG CHIA*

Department of Electronic Engineering, Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, Batu Pahat, Malaysia

**Corresponding author: kschia@uthm.edu.my*

(Received: 17 October 2024; Accepted: 7 November 2024; Published online: 9 September 2025)

ABSTRACT: Accurate and rapid air contaminant classification is crucial for electronic nose (e-nose) systems in air quality monitoring applications. Several one-dimensional convolutional neural networks (1D-CNNs) have recently been proposed for the classification of air contaminants using e-nose systems. However, the lack of cross-model evaluation and the limited computational complexity analysis hinder consistent benchmarking among existing 1D-CNN architectures. Additionally, no recent studies have been conducted on integrating regularization techniques into 1D-CNNs in e-nose. Consequently, the effects of different 1D-CNN architectures, including the impact of integrating regularization techniques, have not been investigated. Thus, this study aims to evaluate three existing 1D-CNN architectures (i.e., LeNet, GasNet, and DenseNet) to propose an improved LeNet with regularization techniques (LeNet-R) for e-nose systems in classifying air contaminants. This study adapted the standard LeNet with three regularization techniques (i.e., batch normalization, dropout, and weight decay) to develop the proposed LeNet-R through a series of manual search experiments. Subsequently, LeNet-R was compared with three existing 1D-CNN models in terms of classification performance and computational complexity using a publicly accessible e-nose dataset. The results show that the proposed LeNet-R outperforms the other 1D-CNN models by achieving the highest average accuracy (i.e., 97.60%) and lowest average loss (i.e., 6.50%). Moreover, LeNet-R exhibited the shortest training time (i.e., 86.54 seconds), the shortest inference time (i.e., 1.91 seconds), the fewest total parameters (i.e., 11,644), and the smallest model size (i.e., 45.48 kB) among all the 1D-CNN models. Compared to the standard LeNet, the proposed LeNet-R improved the average accuracy by 1.35%, reduced total parameters and model size by 11%, shortened training time by 36.6%, and decreased inference time by 6.8%. These findings demonstrate that a simpler 1D-CNN integrated with regularization techniques can outperform more complex 1D-CNN models in classifying air contaminants for an e-nose system. This study is the first to show that integrating three regularization techniques into LeNet can improve accuracy and efficiency for e-nose-based air contaminant classification.

ABSTRAK: Pengelasan bahan pencemar udara yang tepat dan pantas adalah penting bagi sistem hidung elektronik (e-nose) pada aplikasi pemantauan kualiti udara. Kebelakangan ini, beberapa rangkaian neural konvolusi satu dimensi (1D-CNNs) telah dibina bagi tujuan klasifikasi bahan pencemar udara menggunakan sistem e-nose. Walau bagaimanapun, ketiadaan penilaian rentas model serta kekurangan kajian terhadap kerumitan pengiraan telah menyukarkan penanda aras konsisten pada model 1D-CNN sedia ada. Tambahan, tiada kajian terkini mengenai integrasi teknik regularisasi ke atas model 1D-CNN dalam bidang e-nose. Akibatnya, pelbagai senibina 1D-CNN, termasuk impak integrasi teknik regularisasi, belum dapat dikaji dengan sewajarnya. Oleh itu, kajian ini bertujuan menilai tiga senibina 1D-CNN sedia ada (iaitu LeNet, GasNet, dan DenseNet) dengan cadangan penambahbaikan model

LeNet berintegrasikan teknik regularisasi (LeNet-R) untuk sistem e-nose dalam pengelasan bahan pencemar udara. Dalam kajian ini, model LeNet sedia ada, diubah suai dengan tiga teknik regularisasi (iaitu normalisasi kelompok, dropout, dan pereputan berat) bagi membangunkan LeNet-R yang dicadangkan melalui siri eksperimen secara carian manual. Seterusnya, LeNet-R dibandingkan dengan tiga model 1D-CNN sedia ada dari segi prestasi pengelasan serta kerumitan pengiraan menggunakan set data e-nose yang boleh diakses secara umum. Dapatan kajian menunjukkan bahawa LeNet-R mengatasi model 1D-CNN lain dengan mencapai ketepatan pengelasan tertinggi (i.e., 97.60%) dan purata ketidaktepatan terendah (i.e., 6.50%). Tambahan, malalui kaedah LeNet-R masa latihan adalah terpantas (i.e., 86.54 saat), masa inferens paling singkat (i.e., 1.91 saat), jumlah parameter paling sedikit (i.e., 11,644), serta saiz model paling kecil (i.e., 45.48 kB) berbanding model 1D-CNN yang lain. Berbanding LeNet biasa, ketepatan klasifikasi bagi LeNet-R meningkat sebanyak 1.35%, mengurangkan jumlah parameter dan saiz model sebanyak 11%, memendekkan masa latihan sebanyak 36.6%, dan menurunkan masa inferens sebanyak 6.8%. Dapatan menunjukkan bahawa model 1D-CNN yang lebih ringkas dengan integrasi bersama teknik regularisasi mampu mengatasi model 1D-CNN yang lebih kompleks dalam pengelasan bahan pencemar udara untuk sistem e-nose. Kajian ini adalah yang pertama menunjukkan bahawa penyepaduan tiga teknik regularisasi ke dalam LeNet dapat meningkatkan ketepatan dan kecekapan bagi pengelasan pencemar udara berasaskan sistem e-nose.

KEYWORDS: *electronic nose, one-dimensional convolutional neural network, LeNet, air contaminants, classification.*

1. INTRODUCTION

The electronic nose (e-nose) is a device that employs a gas sensor array with a classification algorithm to classify gases [1]. Modeled after the human sense of smell, the e-nose overcomes the limitations of human olfaction and provides a reliable method for identifying a variety of gases. It utilizes a cross-sensitive gas sensor array to capture the properties of target gases, which are then analyzed by a classification algorithm for accurate classification [2]. While e-nose technology has been applied in fields such as medical diagnostics [5] and agriculture [6]. It is particularly valuable for air quality monitoring, where rapid and accurate detection of air contaminants is essential for protecting public health [4]. Harmful air contaminants can negatively impact health and quality of life, contributing to respiratory issues, cardiovascular diseases, and allergies.

Effective air quality monitoring systems that classify various air contaminants are increasingly essential. Traditional methods, such as sensory analysis and physicochemical analysis (e.g., gas chromatography-mass spectrometry [9] and fixed monitoring stations), are often limited by high costs, the need for skilled operators, and a lack of portability, making them impractical for real-time or on-site applications. To address these challenges, e-nose systems have emerged as promising alternatives [8], due to their high accuracy, ease of use, low cost, rapid detection, and ability to provide continuous real-time monitoring [7]. By classifying air contaminants using e-nose systems, improvements in air quality for public health and urban environments can be achieved. This effort supports two Sustainable Development Goal (SDG) targets, i.e., SDG 3 (Good Health and Well-being) and SDG 11 (Sustainable Cities and Communities) [10]. Therefore, e-nose systems are crucial for mitigating potential health risks and creating healthier living environments.

In e-nose systems, achieving accurate air contaminant classification depends on the effectiveness of the classification algorithm [3]. Several traditional machine learning (ML) algorithms have been employed for e-nose-based gas classification, including linear discriminant analysis (LDA) [12], principal component analysis (PCA) [13], and support vector

machine (SVM) [14]. However, these algorithms often require extensive feature engineering and may not fully capture the complex, nonlinear relationships in gas sensor data. To overcome these limitations, deep learning methods, especially convolutional neural networks (CNNs), have been introduced to automate feature extraction and improve classification performance. A recent review study concluded that CNNs can extract features from e-nose data better than ML algorithms [15]. In particular, one-dimensional convolutional neural networks (1D-CNNs) have garnered attention and are being utilized as classification algorithms in air contaminant classification for e-nose systems [16]. The architecture of a 1D-CNN is specifically designed to process one-dimensional (1D) data. The architecture consists of convolutional layers that use 1D filters to extract features by sliding across the input sequence, pooling layers that reduce the data dimensions, and fully connected layers that integrate the extracted features for the final prediction [18]. 1D-CNNs are more suitable classification algorithms than conventional 2D-CNNs due to their lower computational complexity, faster training time, and compact configurations [17]. Studies have demonstrated that 1D-CNNs outperform traditional ML algorithms in extracting features from e-nose data and achieving higher classification accuracy in gas classification tasks [12], [13], [14].

Several existing 1D-CNN models have been proposed for air contaminant classification in e-nose systems. For instance, Peng et al. [19] proposed a novel deep convolutional neural network comprising 38 layers, called GasNet, inspired by combinations of the Residual Network (ResNet) and Visual Geometry Group Network (VGGNet) architectures. Furthermore, Li et al. [20] developed an innovative Densely-connected Convolution Network (DenseNet) with multiple dense blocks and transition layers that utilize 1D convolutions, in contrast to the traditional DenseNet architecture, which uses 2D convolutions. Moreover, Wei et al. [21] proposed an improved LeNet architecture model by modifying the hyperparameters in the convolutional and fully connected layers. Despite the advancements in model architecture, there remains a lack of studies systematically comparing these models using the same e-nose dataset. Moreover, while the benefits of integrating regularization techniques into 1D-CNNs are well established in domains, e.g., phonocardiogram (PCG) [23], polysomnography (PSG) [24], electromyography (EMG) [25], electroencephalogram (EEG) [26], and near-infrared spectroscopy (NIRS) [27], no similar investigations have been reported for e-nose data. Consequently, the effect of different 1D-CNN architectures, including the impact of integrating regularization techniques in e-nose-based air contaminant classification applications, has not been thoroughly investigated.

To address these gaps, this study evaluates the three existing 1D-CNN architectures to propose an improved LeNet with regularization techniques (LeNet-R) for electronic nose (e-nose) systems in classifying air contaminants. The model integrates three regularization techniques, i.e., dropout, batch normalization, and weight decay, into the state-of-the-art LeNet model. This design aims to minimize computational complexity without compromising classification performance. The main contributions of this study are as follows. To the best of our knowledge, this is the first study to integrate three regularization techniques into the LeNet model, which is specifically tailored for e-nose systems in classifying air contaminants. This study also provides a comprehensive evaluation of classification performance and computational complexity, which are critical for applying 1D-CNN models in e-nose technology.

The remainder of this paper is structured as follows: Section 2 describes the methodology and model development. In Section 3, the experimental results are reported and discussed. Lastly, Section 4 outlines key takeaways and directions for future research.

https://doi.org/10.31436/iiumej.v26i3.3471

2. METHODOLOGY

Figure 1 illustrates the flowchart for evaluating the classification performance and computational complexity of each 1D-CNN model in air contaminant classification. A publicly accessible e-nose dataset (Section 2.1) was preprocessed using two methods (Section 2.2), i.e., data partition and data normalization, before being applied to each 1D-CNN model (Section 2.3). Four 1D-CNN models were used in this study, i.e., GasNet, DenseNet, LeNet, and the proposed LeNet-R. All 1D-CNN models were then trained and evaluated to determine the target classes of air contaminants (Section 2.4). Finally, the models were compared in terms of classification performance and computational complexity based on the evaluation metrics (Section 2.5).

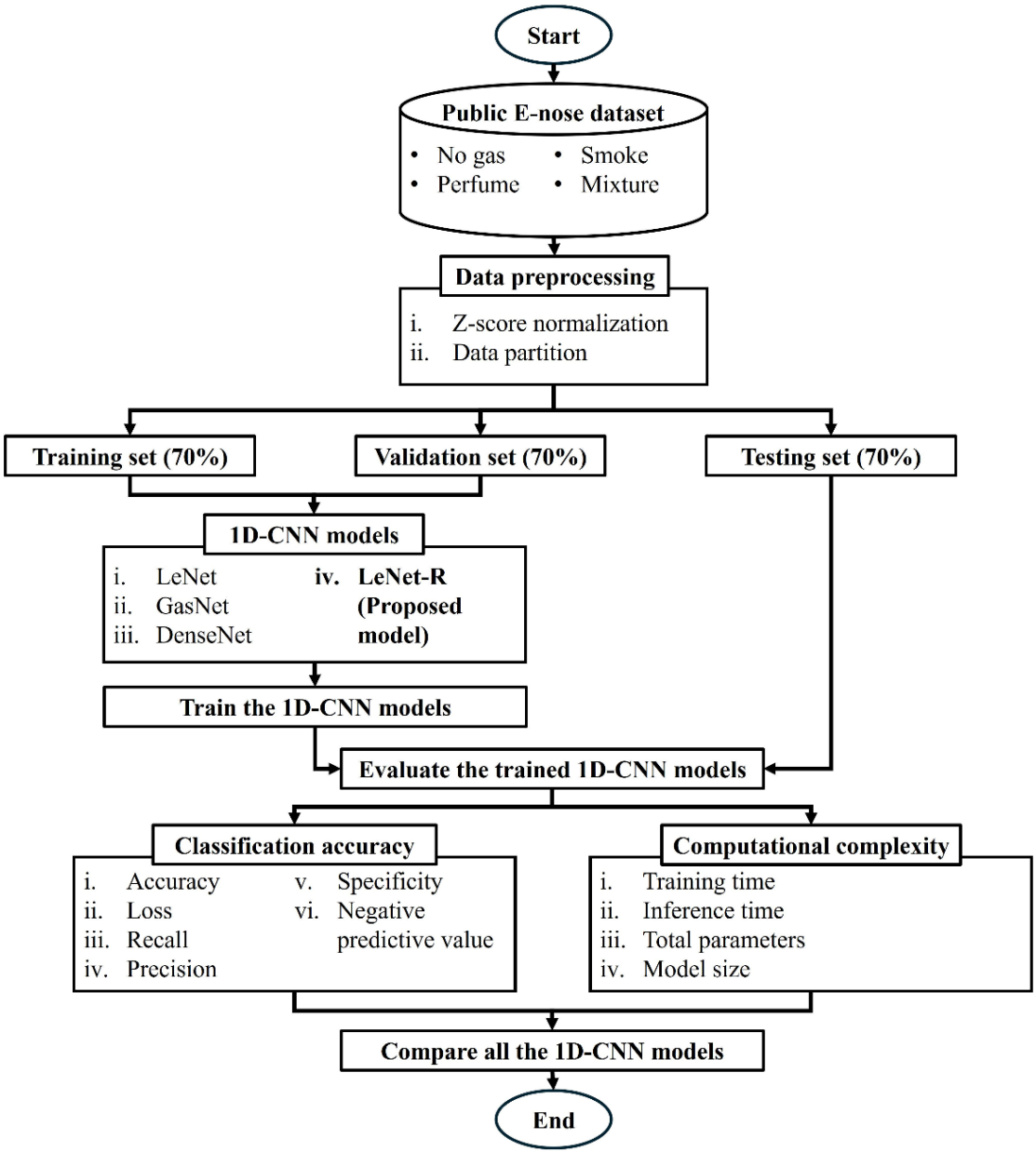


Figure 1. The flowchart for evaluating the classification performance and computational complexity of each 1D-CNN model used in air contaminant classification.

This study was conducted on a laptop computer with an Intel Core i3 CPU running at 2.40 GHz and 8 GB of RAM. The development environment included the frameworks of Keras (version 2.13) and TensorFlow (version 2.14). All 1D-CNN models were trained and evaluated using the R programming language (version 4.3.3) with RStudio software (version 494).

2.1. E-nose dataset

The experimental setup for an e-nose system consists of several key components designed to enable accurate detection and classification of gases, as shown in Figure 2. The system begins with a gas distribution control system that introduces gases into the chamber and is regulated by a mass flow controller, which controls the flow rate of gases to maintain consistent experimental conditions. The core component is the gas sensor array, which detects target gases and outputs signals refined by a signal conditioning circuit. These signals are digitized by a data acquisition system and sent to a computer. Here, a data preprocessing algorithm cleans and prepares the data, while a classification algorithm analyzes the processed signals for accurate gas classification. This setup enables automated gas detection and classification in e-nose applications [28], [29].

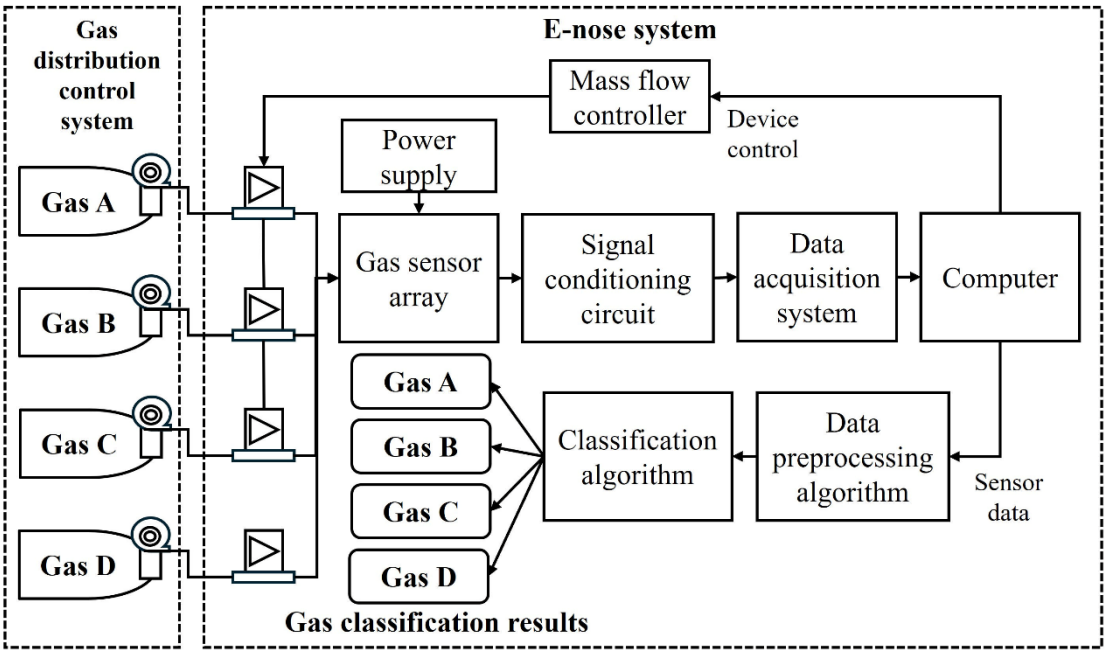


Figure 2. The block diagram of the experimental setup for the e-nose-based gas classification system.

This study used a publicly accessible e-nose dataset that can be obtained from the Mendeley Data repository [28], and the data acquisition methodology is briefly described as follows. This dataset was acquired using seven commercially available gas sensors: MQ2, MQ3, MQ5, MQ6, MQ7, MQ8, and MQ135. Each sensor in this array is sensitive to different gases (e.g., methane, butane, propane, benzine, ethanol, ammonia, liquefied petroleum gas (LPG), alcohol, smoke, natural gas, and carbon monoxide) that enable the detection of a wide variety of gases typically found in industrial and residential environments [30]. Two common gas sources were used: perfume spray and incense stick smoke. These sources were chosen to generate four target classes of air contaminants for the dataset: perfume, smoke, a mixture of perfume and smoke, and no gas (i.e., clean air from the surrounding environment with no

perfume or smoke present). The dataset comprised a total of 6400 samples, with 1600 samples for each class.

This evenly distributed dataset across the four target classes ensures a fair representation of each target class in the dataset. Previous works by Faleh et al. [12] and Narkhede et al. [30], both used 1600 samples per class (with the total of 6400 samples) in their deep learning models and demonstrated robust classification performance. This consistency with prior studies indicates that the dataset size is sufficient for effective training and evaluation of deep learning models, including 1D-CNNs, for air contaminant classification.

2.2. Data preprocessing

Data preprocessing involves two methods: data partition and data normalization. The dataset was divided into a training dataset (70%) and a testing dataset (30%). Then, 20% of the training set was allocated as a validation dataset, which was used to evaluate model performance during the training process continuously. Each target class has a balanced training, validation, and testing dataset. Table 1 shows the sample sizes of the training, validation, and testing datasets for each target class.

All data except for the target variable in the datasets were normalized using the z-score standardization method, where each feature was scaled to achieve a mean of zero and a standard deviation of one [31]. Additionally, the data were reshaped into arrays suitable for model input. The target variable was transformed into a categorical format using one-hot encoding, which is commonly used for multi-class classification problems. Thus, these preprocessing methods ensured the dataset was in a suitable format for an effective 1D-CNN training process.

Table 1. The number of samples for training, validation, and testing datasets of each target class

Target class	Sample Size		
	Training dataset	Validation dataset	Testing dataset
No gas	896	224	480
Perfume	896	224	480
Smoke	896	224	480
Mixture	896	224	480
Total	**3584**	**896**	**1920**

2.3. One-dimensional convolution neural network

In this study, four 1D-CNN models were applied to classify the four target classes of air contaminants. Among these were existing 1D-CNN models (e.g., GasNet, DenseNet, and LeNet), along with the proposed LeNet-R. The existing LeNet, with its simpler architecture, was adapted with three regularization techniques (e.g., batch normalization, weight decay, and dropout) to develop LeNet-R through a series of manual search experiments.

2.3.1. Existing 1D-CNN models

The GasNet consisted of an input layer, six convolutional blocks, two max pooling layers, a global average pooling layer, and a fully connected output layer. The input layer accepted input data with a shape of 7×1. Each convolutional block consisted of two layers, followed by a batch normalization layer and a rectified linear unit (ReLU) activation function layer. The fundamental structure of GasNet is the shortcut connection utilized for convolutional blocks to

https://doi.org/10.31436/iiumej.v26i3.3471

enhance learning capabilities and prevent the vanishing gradient issue [19]. The shortcut connection method, forming a residual block, enhances classification accuracy and smooths training without adding extra parameters or computational load, thereby keeping the layer count constant. The model concluded with a fully connected layer with four units and a Softmax activation function to classify one of the four target air contaminant classes. The hyperparameters and structure of the model were the same as those used in the GasNet architecture [19].

Additionally, the DenseNet consisted of an initial convolutional layer, three dense blocks, two transition layers, a global average pooling layer, and a fully connected output layer. The initial convolutional layer processed the incoming gas sensor data with an input shape of 7×1. DenseNet applied three dense blocks, with the first two blocks each followed by a transition layer, which are the fundamental structures of the model. These dense blocks were designed to enhance the network's feature extraction capabilities, tackle the issue of vanishing gradients, and reduce the total parameter count. The convolutional operations within these blocks utilized a growth rate of 12 to determine the extent of new information that each layer adds to the overall model state. Transition layers played a crucial role in distilling the essence of the input data, enabling an efficient classification process [20]. The compression rate was set to 0.5 to control the number of feature maps reserved for the transition layer. The model concluded with a fully connected layer with four units and a Softmax activation function to classify one of the four target air contaminant classes. The hyperparameters and structure of the model, including dense blocks and transition layers, were the same as those used in the DenseNet architecture [20].

Next, the LeNet consisted of two convolutional layers, two average pooling layers, a flatten layer, and three fully connected layers. The model began with a convolutional layer that employed a ReLU activation function to process incoming gas sensor data, with an input shape of 7×1. Following each convolutional layer, an average pooling layer was applied to reduce the data dimensions of the model. Compared to the Tanh activation function used in the LeNet algorithm, the standard LeNet employed the ReLU activation function in the convolutional and fully connected layers, contributing to faster and more efficient model performance. The model concluded with a fully connected layer with four units and a Softmax activation function to classify one of the four target air contaminant classes. The hyperparameters and structure of the model were the same as those used in the LeNet architecture [21].

2.3.2. The proposed LeNet-R model

The proposed LeNet-R consisted of two convolutional layers, two max pooling layers, a flatten layer, two dropout layers, two batch normalization layers, and three fully connected layers. Table 2 shows the architecture summary of LeNet-R. The first convolutional layer, with eight filters, a kernel size of four, a stride of one, same padding, and a ReLU activation function, processed the incoming gas sensor data with an input shape of 7×1. Following this, a max pooling layer with a pooling size and stride of two was applied to reduce the data dimensions of the model. Batch normalization layers were applied between each convolutional layer and its subsequent max pooling layer. The second convolutional layer, batch normalization layer, and max pooling layer mirrored the first set, except that the number of filters in the convolutional layer was increased to 16. Before transitioning to the fully connected layers, a flatten layer converted the 2D array inputs to a 1D array input. Subsequently, two fully connected layers, each equipped with a ReLU activation function and units of 128 and 64, respectively, processed all inputs from the previous layer. Between each dense layer, a batch normalization layer and a dropout layer with a dropout rate of 0.1 were applied. The model

https://doi.org/10.31436/iiumej.v26i3.3471

concluded with a fully connected layer with four units and a Softmax activation function to classify one of the four target air contaminant classes.

Table 2. The architecture summary of the proposed LeNet-R obtained from R programming.

Layer (type)	Output Shape	Number of Parameters
conv1d_1 (Conv1D)	(32, 7, 8)	40
max_pooling1d_1 (MaxPooling1D)	(32, 3, 8)	0
conv1d_2 (Conv1D)	(32, 3, 8)	528
max_pooling1d_2 (MaxPooling1D)	(32, 1, 16)	0
flatten (Flatten)	(32, 16)	0
dense_1 (Dense)	(32, 128)	2176
batch_normalization_1 (BatchNormalization)	(32, 128)	512
dropout_1 (Dropout)	(32, 128)	0
dense_2 (Dense)	(32, 64)	8256
batch_normalization_2 (BatchNormalization)	(32, 64)	256
dropout_2 (Dropout)	(32, 64)	0
dense_3 (Dense)	(32, 4)	260

The LeNet-R is a classification algorithm whose architecture is adapted from the LeNet architecture. Figure 3 compares the architecture of the standard LeNet and the LeNet-R. The proposed modifications to develop LeNet-R are as follows: the number of filters and the kernel size in the convolutional layers, and the number of units in the fully connected layers, which are set using numbers within a geometric progression [21]. By implementing this method, LeNet-R was able to optimize the CPU and memory resources, resulting in improved classification accuracy and faster model training [32]. This method is also used in deep learning architectures to systematically adjust the computational complexity and model size of neural networks [33]. Moreover, max pooling layers are used in LeNet-R instead of average pooling layers. They capture the maximum value within a data pool, unlike average pooling layers, which take the average value.

Weight decay (L2 regularization) is applied as a kernel regularizer to all convolutional and dense layers, since these layers have learnable parameters. It adds a penalty equal to the sum of the squared values of the model parameters to the loss function, which helps prevent overfitting by encouraging smaller weights [34]. Furthermore, batch normalization layers were applied between the convolutional and max pooling layers to maintain the same parameter distribution and avoid overfitting [20]. In addition, a dropout layer was applied to randomly remove certain connections between units in the following layers during the training process [35].

This study's optimization of hyperparameters for LeNet-R was conducted through a series of intuitive manual search experiments. This optimization specifically addressed the number of filters and kernel sizes in the convolutional layers, the number of units in the fully connected layers, the constant value of bias initialization for both layers, and the dropout rate. The dropout rate was explored within the range of 0.1 to 0.5 to determine the optimal balance between classification performance and avoidance of overfitting. The weight decay coefficient was examined from 0 to 0.01 in increments of 0.001

For the convolutional layers, the approach was to start with fewer filters in the first layer and increase them for the second layer. Conversely, for the fully connected layers, the approach began with more units and reduced the number for the second layer. The number of filters and units was chosen according to a geometric progression, considering the sequential architecture of the model and its impact on learning capability. Additionally, the kernel size for the convolutional layers was explored within the range of one to seven, considering that the dataset

utilizes up to seven sensors. Finally, optimizing hyperparameters is pivotal in defining the scope of the experiment to develop LeNet-R. It ensures the model achieves high classification accuracy with minimal computational complexity for air contaminant classification.

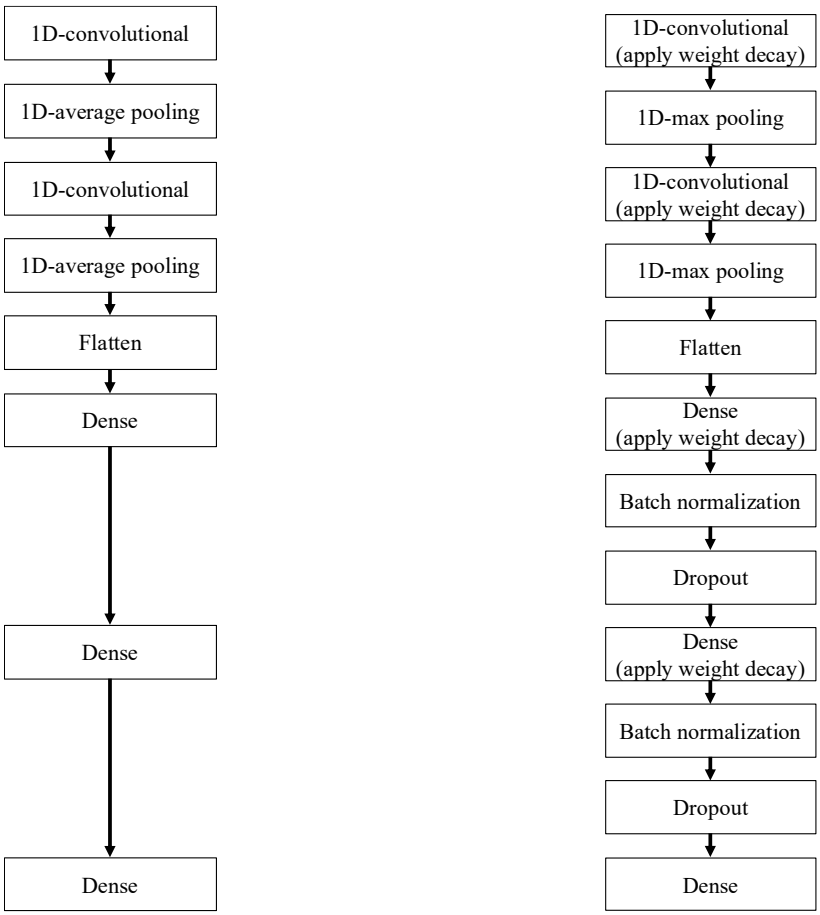


Figure 3. The comparison between the architecture of (a) the standard LeNet and (b) the proposed LeNet-R.

2.4. Model Training

A similar approach was used to train the four 1D-CNN models. This approach ensures that their performance can be compared and evaluated solely based on model architecture differences rather than on training process variations. All the 1D-CNN models were trained in over 100 epochs with a batch size 32. The network updated its weight after processing every 32 samples. Adaptive Moment Estimation (Adam) was used as the optimizer in the network training, with a learning rate set at 0.001. Categorical cross-entropy was applied as the loss function, which is commonly used for multi-class classification problems. To maintain consistency and reproducibility in the results, a random seed number of 42 was set throughout the study [36].

2.5. Evaluation Metrics

The proposed LeNet-R was compared with the three existing 1D-CNN models (i.e. LeNet, DenseNet, and GasNet) in terms of classification performance and computational complexity using a publicly accessible e-nose dataset. The following metrics were obtained to evaluate

https://doi.org/10.31436/iiumej.v26i3.3471

classification performance: accuracy, loss, recall, precision, specificity, and negative predictive value. A confusion matrix was used to assess the classification performance of the 1D-CNN models by categorizing predictions into true positive (TP), false negative (FN), false positive (FP), and true negative (TN). It facilitates the calculation of these metrics by comparing predicted with actual class labels.

Accuracy (Eq. 1) measures the proportion of correct predictions out of all predictions, offering a general view of overall performance. Recall (Eq. 2) focuses specifically on how well the model identifies actual positive cases, while precision (Eq. 3) indicates the proportion of predicted positives that are truly positive. Specificity (Eq. 4) complements recall by measuring the accuracy with which negative cases are classified. Finally, the negative predictive value (Eq. 5) reflects the proportion of predicted negatives that are indeed negative. Together, these metrics comprehensively evaluate classification performance from multiple perspectives. The calculation formulas for these metrics are as follows [16]:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Specificity = \frac{TN}{TN+FP} \quad (4)$$

$$Negative\ predictive\ value = \frac{TN}{TN+FN} \quad (5)$$

Loss refers to the average categorical cross-entropy loss across the training dataset [12]. On the other hand, a learning curve is a graphical tool that charts the progression of learning performance over time, helping to identify any overfitting and underfitting issues. Training time, inference time, total parameters, and model size were considered to evaluate the computational complexity of all 1D-CNN models [37]. Therefore, all these metrics were obtained to present a comprehensive evaluation of the classification performance and computational complexity of each 1D-CNN model.

3. RESULTS AND DISCUSSION

The training performance assessment of the four 1D-CNN models (i.e., GasNet, DenseNet, LeNet, and the proposed LeNet-R) is presented using the learning curve in Section 3.1. Additionally, the evaluation of individual air contaminant classification performance is detailed using the confusion matrix in Section 3.2. Furthermore, the review of all models' average classification performance is discussed in Section 3.3. Section 3.4 presents a comparison of total parameters and model size, and Section 3.5 describes the comparison of training time and inference time for all the models.

3.1. Training performance

The training performance for all 1D-CNN models was observed and evaluated using the learning curves of each model, as shown in Figure 4. It is worth highlighting that all the 1D-CNN models were effectively trained for over 100 epochs. The training and validation datasets were used to train all the 1D-CNN models. The learning curves suggest that all the 1D-CNN models achieved an optimal fit, with no evidence of underfitting or overfitting issues during network training.

https://doi.org/10.31436/iiumej.v26i3.3471

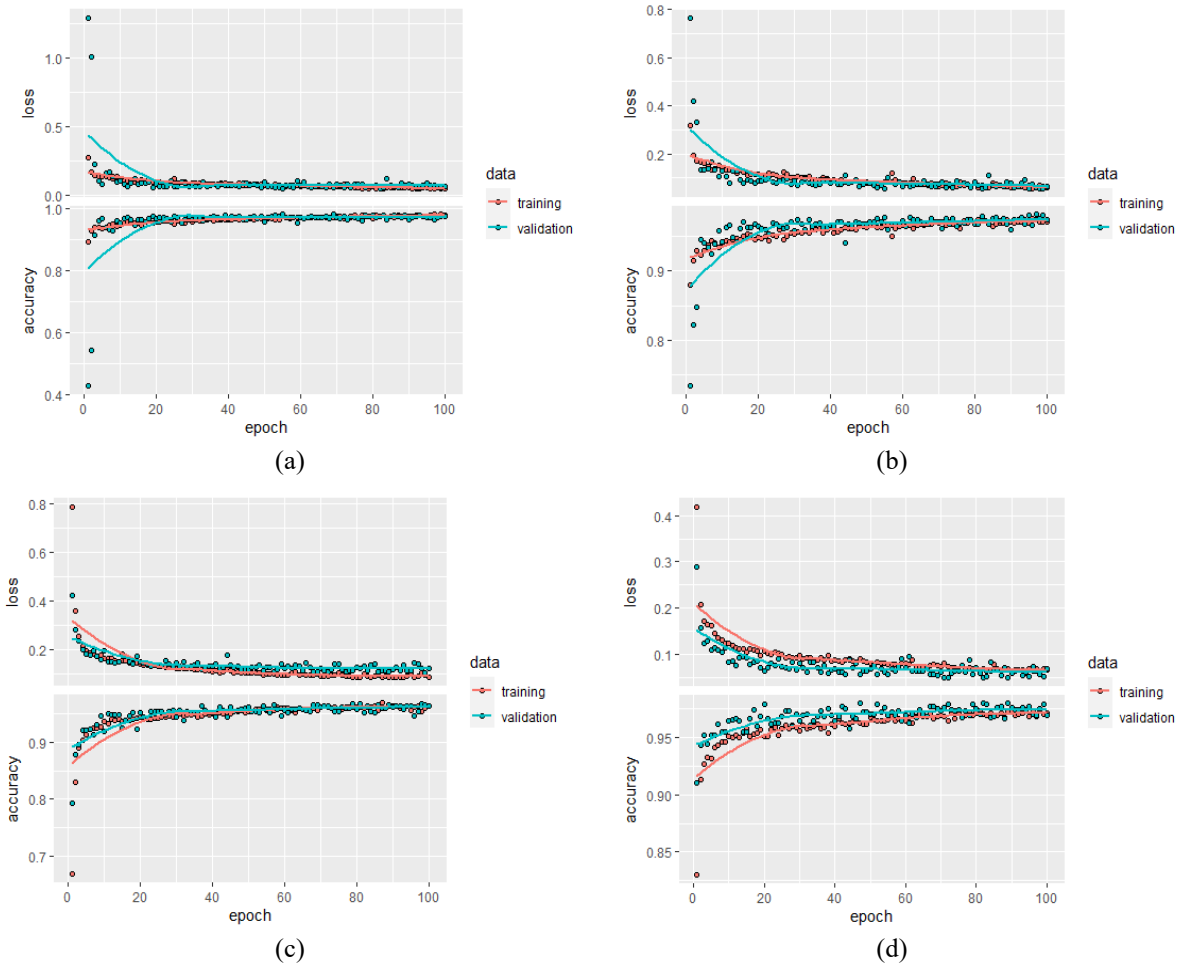


Figure 4. Training and validation learning curves for the four 1D-CNN models: (a) GasNet, (b) DenseNet, (c) LeNet, and (d) the proposed LeNet-R.

3.2. Individual air contaminant classification performance

The classification performance for all the 1D-CNN models was evaluated using the confusion matrix corresponding to each model, as illustrated in Figure 5. The testing dataset was used for the confusion matrix to assess the classification performance of all the models. The confusion matrices demonstrate that all the models are highly effective in accurately classifying two target air contaminants, i.e., smoke and a mixture of smoke and perfume, with no false positives or false negatives detected. However, the two other target classes of air contaminants, i.e., perfume and no gas from the surrounding environment, exhibited a few instances of false positives and false negatives.

Among all models, the proposed LeNet-R exhibited the lowest total number of misclassified samples (i.e., 46), followed by GasNet (i.e., 61), DenseNet (i.e., 66), and LeNet (i.e., 71). This further demonstrates the effectiveness of the proposed LeNet-R model in accurately distinguishing between the different air contaminant classes, particularly in reducing confusion between no gas and perfume. Derived from each confusion matrix, the key metrics (i.e., recall, precision, specificity, and negative predictive value) for each target class were calculated for all the models, as detailed in Table 3.

https://doi.org/10.31436/iiumej.v26i3.3471

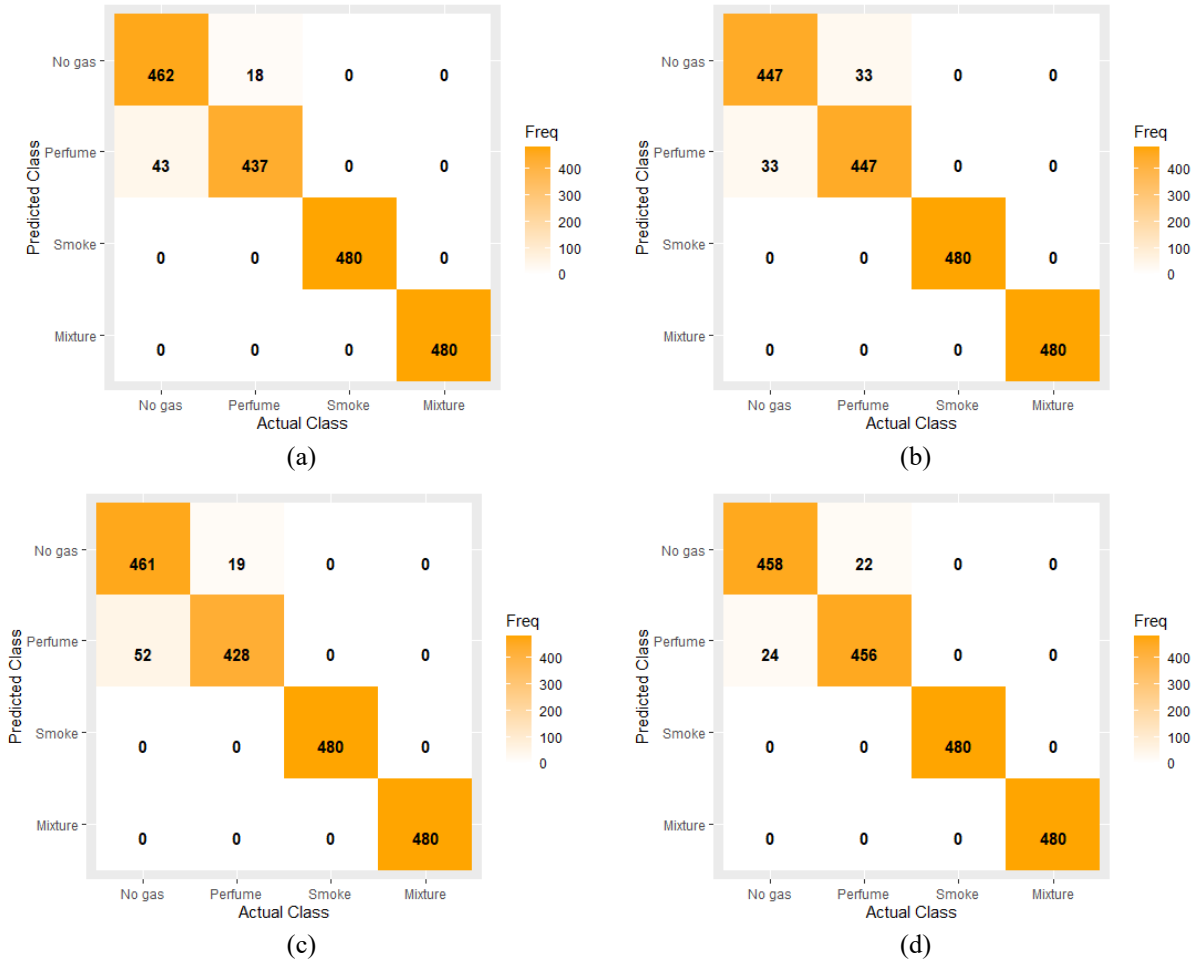


Figure 5. The confusion matrices of (a) GasNet, (b) DenseNet, (c) LeNet, and (d) the proposed LeNet-R.

Table 3. The recall, precision, specificity, and negative predictive value metrics for all four 1D-CNN models across each target class.

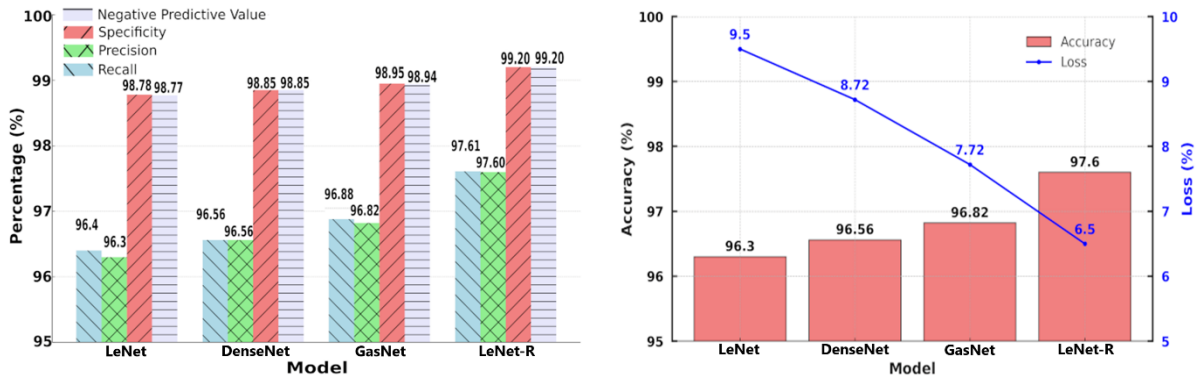
Model	Class	Accuracy (%)	Recalling (%)	Precision (%)	Specificity (%)	Negative Predictive Value (%)
LeNet [21]	No gas	89.86	89.86	96.04	**98.65**	96.39
	Perfume	95.75	95.75	89.17	96.47	98.68
	Smoke	100.00	100.00	100.00	100.00	100.00
	Mixture	100.00	100.00	100.00	100.00	100.00
DenseNet [20]	No gas	93.13	93.13	93.13	97.71	97.71
	Perfume	93.13	93.13	93.13	97.71	97.71
	Smoke	100.00	100.00	100.00	100.00	100.00
	Mixture	100.00	100.00	100.00	100.00	100.00
GasNet [19]	No gas	91.49	91.49	**96.25**	98.73	97.01
	Perfume	**96.04**	**96.04**	91.04	97.07	**98.75**
	Smoke	100.00	100.00	100.00	100.00	100.00
	Mixture	100.00	100.00	100.00	100.00	100.00
LeNet-R (Proposed)	No gas	**95.02**	**95.02**	95.42	98.47	**98.33**
	Perfume	95.40	95.40	**95.00**	**98.34**	98.47
	Smoke	100.00	100.00	100.00	100.00	100.00
	Mixture	100.00	100.00	100.00	100.00	100.00

A detailed analysis of the classification metrics in Table 3 shows that the proposed LeNet-R model achieves the highest number of optimum performance metrics (i.e., 5) among all evaluated models. Specifically, LeNet-R demonstrates the highest results in accuracy, recall, and negative predictive value for the no gas class and precision and specificity for the perfume class. GasNet achieves four optimum performance metrics, including the highest precision for the 'no gas' class and the highest accuracy, recall, and negative predictive value for the 'perfume' class. In contrast, the standard LeNet achieves only one optimum performance metric, i.e., the highest specificity for no gas, while DenseNet does not lead in any key metric. These results highlight the effectiveness of the proposed LeNet-R in reducing misclassification for the more challenging 'no gas' and 'perfume' classes.

3.3. Average classification performance evaluation

Figure 6(a) presents the classification performance evaluation of all 1D-CNN models, focusing on the average recall, precision, specificity, and negative predictive value for each target class obtained by each model. The proposed LeNet-R achieved the highest recall, precision, specificity, and negative predictive value of 97.61, 97.60, 99.20, and 99.20%, respectively, followed by GasNet, DenseNet, and LeNet. The high recall of the LeNet-R model significantly reduces the risk of misclassifying the intended target class as another. Similarly, the high precision minimizes the probability of incorrectly predicting the target class of air contaminants. Additionally, the high specificity plays a crucial role in accurately classifying the non-target class of air contaminants. Furthermore, the high negative predictive value provides high confidence in confirming the absence of any target classes of air contaminants. Thus, these metrics show the capabilities of the LeNet-R as the most robust and reliable model for e-nose applications, particularly air quality monitoring, by classifying air contaminants.

Figure 6(b) presents the classification performance evaluation of all 1D-CNN models in terms of averaged accuracy and averaged loss. The proposed LeNet-R outperformed all other models with the highest averaged accuracy of 97.60% and the lowest averaged loss of 6.50%, followed by GasNet, DenseNet, and LeNet. The high accuracy of the proposed LeNet-R underscores its superior reliability in correctly classifying different air contaminants. Additionally, its low loss indicates fewer mistakes during the training process in classifying air contaminants. Thus, the proposed modifications to develop the LeNet-R contribute to its higher classification performance across all key metrics than the other models.



(a) (b)

Figure 6. The average classification performance evaluation for all four 1D-CNN models: (a) averaged recall, averaged precision, averaged specificity, and averaged negative predictive value, and (b) averaged accuracy and averaged loss.

Table 4 presents a benchmarking comparison of the averaged accuracy between previous studies and the models evaluated in this study using the same e-nose dataset. Notably, the proposed LeNet-R achieved the highest average accuracy of 97.60%, compared to 96.30% for the standard LeNet model. This corresponds to a relative improvement of 1.35% in average accuracy. It is worth highlighting previous works that used 1D-CNN [12] and CNN [30] with the same dataset and training-to-testing ratio, the testing accuracy of 90% and 93%, respectively. These results demonstrate that the proposed LeNet-R is the most effective model for air contaminant classification using this benchmark e-nose dataset.

Table 4. The average accuracy of the proposed LeNet-R and other models using the same e-nose dataset

Study	Model	Averaged Accuracy
Previous study	1D-CNN [12]	90.00%
	CNN [30]	93.00%
This study	LeNet	96.30%
	DenseNet	96.56%
	GasNet	96.82%
	LeNet-R (Proposed model)	97.60%

3.4. Total parameters and model size

Figure 7(a) compares computational complexity for all 1D-CNN models in terms of total parameters and model size. The proposed LeNet-R exhibited the lowest total parameters of 11,644 and the smallest compact model size of 45.48 kilobytes (kB), followed by LeNet, DenseNet, and GasNet. Specifically, compared to the standard LeNet, LeNet-R reduces total parameters and model size by approximately 10.9% and 11.0%, respectively.

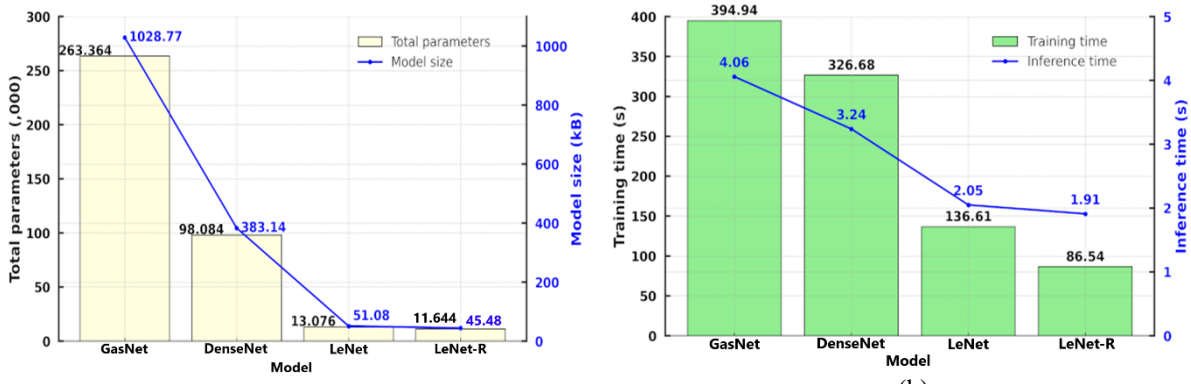


Figure 7. The comparison of computational complexity for all four 1D-CNN models: (a) total parameters and model size, and (b) training time and inference time.

These metrics (i.e., total parameters and model size) help understand each model's practicality for an e-nose system. Models with fewer parameters are advantageous because they require less memory, offer faster training and inference times, and are more suitable for deployment on resource-constrained devices. Smaller models also help conserve battery life and improve overall device performance by minimizing memory usage and computational demand [38]. The proposed adaptations to develop the LeNet-R contribute to a reduction in the total parameters and model size without compromising the classification performance.

Thus, the LeNet-R establishes itself as the most reliable model for air contaminant classification in an e-nose system compared to the other three existing 1D-CNN models.

3.5. Training time and inference time

Figure 7(b) compares computational complexity in terms of training time and inference time. The proposed LeNet-R outperformed all other models with the fastest training time and inference time at 86.54 and 1.91 seconds, respectively, followed by LeNet, DenseNet, and GasNet. Compared to LeNet, LeNet-R reduced training time by 36.6% and inference time by 6.8%, while still achieving superior classification accuracy. The LeNet-R stands out for its ability to provide the fastest classification results without significantly compromising classification performance. Thus, the LeNet-R establishes itself as the most reliable model for the e-nose application, where speed and efficiency are critical. The speed and efficiency of the model are particularly beneficial for e-nose systems requiring immediate data analysis and decision-making.

The results indicate that models with larger total parameters and model size, due to the depth and complexity of the network architecture, i.e., GasNet and DenseNet, require longer training time and inference time than simpler models, i.e., LeNet and LeNet-R. However, the LeNet-R achieved faster training time and inference time, lower total parameters, and a smaller model size than LeNet due to the smaller kernel size in convolutional layers and fewer units in fully connected layers. Rapid identification is as crucial as high accuracy, since delays in detection can increase health risks. This balance of effectiveness and efficiency of a classification algorithm is crucial for real-world e-nose applications. High accuracy ensures reliable detection of hazardous gases, while a smaller and faster model enables rapid, energy-efficient operation on portable e-nose devices [39]. Therefore, the proposed LeNet-R proves to be the simplest 1D-CNN model with the highest classification accuracy for an e-nose system in air contaminant classification compared to the other three existing 1D-CNN models.

4. CONCLUSION

An improved LeNet with regularization techniques (LeNet-R) for air contaminant classification was proposed in this study. The proposed LeNet-R was developed by adapting the LeNet architecture with three regularization techniques (i.e., batch normalization, dropout, and weight decay) through a series of manual search experiments. The proposed LeNet-R was compared with the three existing 1D-CNN models (i.e., LeNet, DenseNet, and GasNet) regarding classification performance and computational complexity using a publicly accessible e-nose dataset. Four target air contaminants were considered: perfume, smoke, a mixture of perfume and smoke, and no gas (i.e., clean air from the surrounding environment with no perfume or smoke present). Each target class had a balanced training, validation, and testing dataset. The findings reveal that the proposed LeNet-R outperforms the three 1D-CNN models by achieving the highest averaged accuracy (i.e., 97.60%) and the lowest averaged loss (i.e., 6.50%). Additionally, the proposed LeNet-R achieved the highest recall, precision, specificity, and negative predictive value of 97.61, 97.60, 99.20, and 99.20%, respectively, followed by GasNet, DenseNet, and LeNet. Furthermore, the LeNet-R produced the shortest training time (i.e., 86.54 seconds), the shortest inference time (i.e., 1.91 seconds), the fewest total parameters (i.e., 11,644), and the smallest model size (i.e., 45.48 kB), followed by LeNet, DenseNet, and GasNet. Thus, the proposed adaptations to develop the LeNet-R contribute to reducing the total parameters and model size without compromising the classification performance. The proposed LeNet-R is the simplest 1D-CNN model with the highest classification accuracy for an e-nose system in air contaminant classification compared to the three existing 1D-CNN

models. Therefore, this study shows that a simpler 1D-CNN integrated with regularization techniques can outperform more complex 1D-CNN models in classifying air contaminants for an e-nose system. In future work, the proposed LeNet-R will be enhanced using hyperparameter tuning algorithms to increase classification accuracy further.

ACKNOWLEDGEMENT

This research was supported by the Universiti Tun Hussein Onn Malaysia (UTHM) through Multi Disciplinary Research Grant (MDR) (vot Q773), and Geran Penyelidikan Pascasiswazah (GPPS) (vot Q570).

REFERENCES

[1] S. Kr. Jha, R. D. S. Yadava, K. Hayashi, and N. Patel, "Recognition and sensing of organic compounds using analytical methods, chemical sensors, and pattern recognition approaches," Chemometrics and Intelligent Laboratory Systems, vol. 185, pp. 18–31, Feb. 2019, doi: 10.1016/j.chemolab.2018.12.008.

[2] Y. Shi, B. Wang, C. Yin, Z. Li, and Y. Yu, "Performance improvement: A lightweight gas information classification method combined with an electronic nose system," Sens Actuators B Chem, vol. 396, p. 134551, Dec. 2023, doi: 10.1016/j.snb.2023.134551.

[3] S. Zhai, Z. Li, H. Zhang, L. Wang, S. Duan, and J. Yan, "A multilevel interleaved group attention-based convolutional network for gas detection via an electronic nose system," Eng Appl Artif Intell, vol. 133, p. 108038, Jul. 2024, doi: 10.1016/j.engappai.2024.108038.

[4] J. Oh et al., "Machine learning-based discrimination of indoor pollutants using an oxide gas sensor array: High endurance against ambient humidity and temperature," Sens Actuators B Chem, vol. 364, p. 131894, Aug. 2022, doi: 10.1016/j.snb.2022.131894.

[5] D. Aulia, R. Sarno, S. C. Hidayati, and M. Rivai, "Optimization of the Electronic Nose Sensor Array for Asthma Detection Based on Genetic Algorithm," IEEE Access, vol. 11, pp. 74924–74935, 2023, doi: 10.1109/ACCESS.2023.3291451.

[6] H. Sun, Z. Hua, C. Yin, F. Li, and Y. Shi, "Geographical traceability of soybean: An electronic nose coupled with an effective deep learning method," Food Chem, vol. 440, May 2024, doi: 10.1016/j.foodchem.2023.138207.

[7] H. Chen, D. Huo, and J. Zhang, "Gas Recognition in E-Nose System: A Review," IEEE Trans Biomed Circuits Syst, vol. 16, no. 2, pp. 169–184, Apr. 2022, doi: 10.1109/TBCAS.2022.3166530.

[8] N. Kar Chowdhury, A. Kumar Singh, and B. Bhowmik, "Role of gas sensor an integrated part of e-nose for online monitoring of air grade," in Nanotechnology-Based E-noses, Elsevier, 2023, pp. 377–394. doi: 10.1016/B978-0-323-91157-3.00005-2.

[9] M. Moufid, B. Bouchikhi, C. Tiebe, M. Bartholmai, and N. El Bari, "Assessment of outdoor odor emissions from polluted sites using simultaneous thermal desorption-gas chromatography-mass spectrometry (TD-GC-MS), electronic nose in conjunction with advanced multivariate statistical approaches," Atmos Environ, vol. 256, p. 118449, Jul. 2021, doi: 10.1016/j.atmosenv.2021.118449.

[10] T. H. Son, Z. Weedon, T. Yigitcanlar, T. Sanchez, J. M. Corchado, and R. Mehmood, "Algorithmic urban planning for smart and sustainable development: Systematic review of the literature," Sustain Cities Soc, vol. 94, p. 104562, Jul. 2023, doi: 10.1016/j.scs.2023.104562.

[11] V. Saiu, I. Blečić, and I. Meloni, "Making sustainability development goals (SDGs) operational at suburban level: Potentials and limitations of neighbourhood sustainability assessment tools," Environ Impact Assess Rev, vol. 96, p. 106845, Sep. 2022, doi: 10.1016/j.eiar.2022.106845.

[12] R. Faleh and A. Kachouri, "A hybrid deep convolutional neural network-based electronic nose for pollution detection purposes," Chemometrics and Intelligent Laboratory Systems, vol. 237, p. 104825, Jun. 2023, doi: 10.1016/j.chemolab.2023.104825.

[13] Y. Li, Z. Ye, and Q. Li, "Precise Identification of Food Smells to Enable Human–Computer Interface for Digital Smells," Electronics (Basel), vol. 12, no. 2, p. 418, Jan. 2023, doi: 10.3390/electronics12020418.

[14] [14] D. Yang, N. Hou, J. Lu, and D. Ji, "Novel leakage detection by ensemble 1DCNN-VAPSO-SVM in oil and gas pipeline systems," Appl Soft Comput, vol. 115, p. 108212, Jan. 2022, doi: 10.1016/j.asoc.2021.108212.

[15] Z. Ye, Y. Liu, and Q. Li, "Recent Progress in Smart Electronic Nose Technologies Enabled with Machine Learning Methods," Sensors, vol. 21, no. 22, p. 7620, Nov. 2021, doi: 10.3390/s21227620.

[16] X. Yang, M. Li, X. Ji, J. Chang, Z. Deng, and G. Meng, "Recognition Algorithms in E-Nose: A Review," IEEE Sens J, vol. 23, no. 18, pp. 20460–20472, Sep. 2023, doi: 10.1109/JSEN.2023.3302868.

[17] J. Kim, S. Oh, H. Kim, and W. Choi, "Tutorial on time series prediction using 1D-CNN and BiLSTM: A case example of peak electricity demand and system marginal price prediction," Eng Appl Artif Intell, vol. 126, p. 106817, Nov. 2023, doi: 10.1016/j.engappai.2023.106817.

[18] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," Mech Syst Signal Process, vol. 151, Apr. 2021, doi: 10.1016/j.ymssp.2020.107398.

[19] P. Peng, X. Zhao, X. Pan, and W. Ye, "Gas Classification Using Deep Convolutional Neural Networks," Sensors, vol. 18, no. 2, p. 157, Jan. 2018, doi: 10.3390/s18010157.

[20] F. Li et al., "A novel DenseNet with warm restarts for gas recognition in complex airflow environments," Microchemical Journal, vol. 197, p. 109864, Feb. 2024, doi: 10.1016/j.microc.2023.109864.

[21] G. Wei, G. Li, J. Zhao, and A. He, "Development of a LeNet-5 Gas Identification CNN Structure for Electronic Noses," Sensors, vol. 19, no. 1, p. 217, Jan. 2019, doi: 10.3390/s19010217.

[22] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998, doi: 10.1109/5.726791.

[23] M. Morshed and S. A. Fattah, "SAR-CardioNet: A Network for Heart Valve Disease Detection From PCG Signal Based on Split-Self Attention With Residual Paths," IEEE Sens J, vol. 23, no. 17, pp. 19553–19560, Sep. 2023, doi: 10.1109/JSEN.2023.3289109.

[24] S. Sanghavi, P. Vaid, P. Rathod, and K. Srivastava, "SpectroTemporalNet: Automated Sleep Stage Scoring with Stacked Generalization," in 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), IEEE, Aug. 2021, pp. 1256–1263. doi: 10.1109/ICESC51422.2021.9532640.

[25] Y. Zhang et al., "EMG-Based Cross-Subject Silent Speech Recognition Using Conditional Domain Adversarial Network," IEEE Trans Cogn Dev Syst, vol. 15, no. 4, pp. 2282–2290, Dec. 2023, doi: 10.1109/TCDS.2023.3316701.

[26] H. Mukhtar, S. M. Qaisar, and A. Zaguia, "Deep Convolutional Neural Network Regularization for Alcoholism Detection Using EEG Signals," Sensors, vol. 21, no. 16, p. 5456, Aug. 2021, doi: 10.3390/s21165456.

[27] [27] Y. Yu, J. Huang, L. Wang, and S. Liang, "A 1D-inception-ResNet based global detection model for thin-skinned multifruit spectral quantitative analysis," Food Control, vol. 167, p. 110823, Jan. 2025, doi: 10.1016/j.foodcont.2024.110823.

[28] [28] P. Narkhede, R. Walambe, P. Chandel, S. Mandaokar, and K. Kotecha, "MultimodalGasData: Multimodal Dataset for Gas Detection and Classification," Data (Basel), vol. 7, no. 8, p. 112, Aug. 2022, doi: 10.3390/data7080112.

[29] J. Fonollosa, L. Fernández, A. Gutiérrez-Gálvez, R. Huerta, and S. Marco, "Calibration transfer and drift counteraction in chemical sensor arrays using Direct Standardization," Sens Actuators B Chem, vol. 236, pp. 1044–1053, Nov. 2016, doi: 10.1016/j.snb.2016.05.089.

[30] P. Narkhede, R. Walambe, S. Mandaokar, P. Chandel, K. Kotecha, and G. Ghinea, "Gas Detection and Identification Using Multimodal Artificial Intelligence Based Sensor Fusion," Applied System Innovation, vol. 4, no. 1, p. 3, Jan. 2021, doi: 10.3390/asi4010003.

[31] Md. Z. Uddin and M. M. Hassan, "Activity Recognition for Cognitive Assistance Using Body Sensors Data and Deep Convolutional Neural Network," IEEE Sens J, vol. 19, no. 19, pp. 8413–8419, Oct. 2019, doi: 10.1109/JSEN.2018.2871203.

[32] Y. An, X. Ma, X. Wang, Z. Qu, X. Zhu, and W. Yin, "Gas pipeline event classification based on one-dimensional convolutional neural network," Struct Health Monit, vol. 21, no. 3, pp. 826–834, May 2022, doi: 10.1177/14759217211010270.

[33] V. Pareek and S. Chaudhury, "Deep learning-based gas identification and quantification with auto-tuning of hyper-parameters," Soft comput, vol. 25, no. 22, pp. 14155–14170, Nov. 2021, doi: 10.1007/s00500-021-06222-1.

[34] H. Mohr, U. Wolfensteller, S. Frimmel, and H. Ruge, "Sparse regularization techniques provide novel insights into outcome integration processes," Neuroimage, vol. 104, pp. 163–176, Jan. 2015, doi: 10.1016/j.neuroimage.2014.10.025.

[35] D. Passos and P. Mishra, "A tutorial on automatic hyperparameter tuning of deep spectral modelling for regression and classification tasks," Chemometrics and Intelligent Laboratory Systems, vol. 223, p. 104520, Apr. 2022, doi: 10.1016/j.chemolab.2022.104520.

[36] K. Kaczmarczyk and K. Miałkowska, "Backtesting comparison of machine learning algorithms with different random seed," Procedia Comput Sci, vol. 207, pp. 1901–1910, 2022, doi: 10.1016/j.procs.2022.09.248.

[37] E. Koşar and B. Barshan, "A new CNN-LSTM architecture for activity recognition employing wearable motion sensor data: Enabling diverse feature extraction," Eng Appl Artif Intell, vol. 124, p. 106529, Sep. 2023, doi: 10.1016/j.engappai.2023.106529.

[38] C. Garbin, X. Zhu, and O. Marques, "Dropout vs. batch normalization: an empirical study of their impact to deep learning," Multimed Tools Appl, vol. 79, no. 19–20, pp. 12777–12815, May 2020, doi: 10.1007/s11042-019-08453-9.

[39] L. Wang, F. Li, C. Yang, L. Feng, and X. Cao, "Performance evaluation of lightweight pattern recognition algorithms for portable environmental monitoring electronic noses," Build Environ, vol. 269, p. 112446, Feb. 2025, doi: 10.1016/j.buildenv.2024.112446.

https://doi.org/10.31436/iiumej.v26i3.3610

PERFORMANCE BENCHMARKING OF HYPERLEDGER FABRIC ON HETEROGENEOUS HARDWARE FOR IOT APPLICATIONS

MUHAMMAD MUAZ ZULKARNAIN[1], NABILAH RAMLI[1*], ANIS NURASHIKIN NORDIN[2]

[1]*Department of Mechanical and Aerospace Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia*
[2]*Department of Electrical and Computer Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia*

**Corresponding author: nabilah@iium.edu.my*

(Received: 13 February 2025; Accepted: 22 August 2025; Published online: 9 September 2025)

ABSTRACT: Hyperledger Fabric (HLF), a widely used open-source private blockchain, has garnered attention for its ability to enable blockchain in Internet of Things (IoT) applications. Embedding HLF nodes within IoT devices enables smart contract integration for secure and automated communications, thereby reducing reliance on intermediaries. However, resource-constrained IoT devices often face challenges with complex operations due to their limited processing power. While HLF deployment on single-board computers (SBCs) like Raspberry Pi has been explored, comprehensive performance evaluations across diverse hardware setups in a heterogeneous blockchain network are limited. This study benchmarks HLF performance on a network comprising SBCs with ARM architectures (Cortex-A72 and Cortex-A76) and laptops with Intel Core i7 and Intel Celeron processors. Using Hyperledger Caliper, key performance metrics, including throughput, latency, CPU usage, and memory consumption, were measured. Results show that high-throughput applications are best supported by high-end processors capable of handling multiple clients, achieving up to 1,148.3 TPS. In contrast, SBCs efficiently handle moderate transaction loads from single clients with minimal latency increases. These findings demonstrate the adaptability of HLF across varied hardware configurations, provided a proper network architecture setup, supporting its deployment in diverse IoT environments.

ABSTRAK: Fabrik hiperlejer (Hyperledger Fabric, HLF), iaitu rantaian blok peribadi sumber terbuka yang digunakan secara meluas, telah mendapat perhatian kerana keupayaannya dalam memacu blok rantaian dalam aplikasi Internet of Things (IoT). Penyepaduan nod HLF ke dalam peranti IoT membolehkan penggunaan kontrak pintar bagi komunikasi selamat dan automatik, seterusnya mengurangkan kebergantungan kepada pihak ketiga. Namun, peranti IoT yang memiliki sumber terhad sering menghadapi cabaran dalam melaksanakan operasi kompleks disebabkan oleh kuasa pemprosesannya yang terhad. Walaupun penerapan nod HLF pada komputer papan tunggal (SBC) seperti Raspberry Pi telah dikaji, penilaian prestasi yang komprehensif merangkumi pelbagai konfigurasi perkakasan dalam blok rantaian heterogen masih terhad. Kajian ini merupakan penanda aras prestasi HLF pada rangkaian yang terdiri daripada SBC berasaskan seni bina ARM (Cortex-A72 dan Cortex-A76) serta komputer riba berprosesor Intel Core i7 dan Intel Celeron. Mengguna pakai Hyperledger Caliper, metrik prestasi utama seperti kadar penghantaran, masa tindak balas, penggunaan CPU, dan penggunaan memori telah diukur. Dapatan kajian menunjukkan bahawa kadar aplikasi penghantaran tertinggi, paling sesuai disokong oleh pemproses berprestasi tinggi yang mampu mengendalikan pelbagai klien, mencapai sehingga 1,148.3 TPS. Sementara itu, SBC berupaya mengendalikan beban transaksi sederhana daripada klien tunggal dengan peningkatan masa tindak balas yang minima. Penemuan ini menunjukkan kebolehsesuaian

https://doi.org/10.31436/iiumej.v26i3.3610

HLF merentasi pelbagai konfigurasi perkakasan, dengan syarat rangkaian yang sesuai disediakan, sekaligus menyokong pelaksanaannya dalam persekitaran IoT yang pelbagai.

KEY WORDS: *Hyperledger Fabric, Hyperledger Caliper, Internet of Things (IoT), Performance Analysis, IoT-Enabled Blockchain*

1. INTRODUCTION

The rapid evolution of the IoT has transformed industries by enabling interconnected devices to process and exchange data in real-time, playing a pivotal role in sustainability and environmental management [1]. IoT drives advancements in renewable energy adoption, energy efficiency, and carbon offset tracking, aligning with global efforts to combat environmental challenges and achieve the Sustainable Development Goals (SDGs) [2], [3]. When integrated with blockchain technology, IoT's capabilities are further enhanced through decentralized, secure, and immutable ledgers that enable applications such as peer-to-peer energy trading, real-time energy monitoring, and carbon credit traceability [4], [5]. These innovations strengthen data security, improve traceability, and promote sustainable energy practices, accelerating the shift toward cleaner energy systems [6].

However, traditional centralized IoT systems often suffer from single points of failure, latency issues, and cybersecurity risks [7], [8]. Blockchain technology addresses these problems by decentralizing data management and enhancing security. Integrating blockchain directly into IoT devices as single units with in-situ capability offers the potential for improved efficiency and autonomy. Despite these advantages, this approach introduces new challenges, particularly scalability concerns and the need for efficient network coordination in distributed environments.

A key challenge in integrating blockchain with IoT devices in situ lies in the resource constraints of IoT devices. These devices frequently operate with limited CPU, memory, and storage capacity, which are further strained by blockchain operations such as cryptographic signing, state database updates, and consensus participation [9]. Furthermore, consensus mechanisms such as proof of work (PoW) further exacerbate these issues, introducing high latency and scalability challenges, especially in large-scale deployments [10].

In heterogeneous IoT networks, hardware performance varies significantly — from high-end servers or laptops acting as gateways to low-power single-board computers (SBCs) operating at the edge. Understanding how these different classes of devices perform under blockchain workloads is critical for IoT architects, as it informs role assignment, workload balancing, and cost-performance trade-offs.

This paper aims to benchmark Hyperledger Fabric performance across heterogeneous IoT-representative hardware (high-end, mid-range, and SBC) by investigating a system network architecture using Hyperledger Fabric deployed on heterogeneous hardware, including two personal computers with Intel Core i7-1355U @ 1.70 GHz and Intel Celeron 3205U @ 1.50 GHz processors, and two SBCs with ARM Cortex-A72 @ 1.8 GHz and Cortex-A76 @ 2.4 GHz processors. This combination reflects real-world IoT deployment scenarios where central nodes possess high processing power, while distributed edge devices are cost-efficient and energy-efficient but resource-limited. Performance was benchmarked using Hyperledger Caliper (HLC) [11] to evaluate key metrics such as transaction throughput, latency, and resource utilization to analyze the impact of hardware specifications in distributed blockchain-based IoT ecosystems. By directly linking these performance metrics to IoT operational requirements, the study provides insights into how hardware constraints affect blockchain

https://doi.org/10.31436/iiumej.v26i3.3610

performance in scenarios such as carbon offset reporting, energy metering, and other data-intensive IoT services.

The paper is organized as follows: Section II covers the background knowledge on the Hyperledger system and related works; Section III outlines the methodology approach; Section IV presents the results and discusses the system's performance based on the key metrics. Finally, Section V concludes the paper by summarizing the findings and suggesting directions for future work.

2. BACKGROUND

2.1. Hyperledger Fabric

Hyperledger Fabric is a permissioned blockchain platform for enterprise-grade applications, offering a modular architecture for flexible customization of components such as consensus protocols and membership services [12]. Hyperledger Fabric supports lightweight consensus mechanisms like Raft, making it highly efficient and well-suited for applications requiring high throughput and low resource consumption. As illustrated in Figure 1, the architecture highlights the core component of Hyperledger Fabric, the Certificate Authority (CA), which is responsible for issuing digital certificates to verify and authenticate network participants, ensuring that only authorized entities can access and interact with the network. Another key element is the peer, which maintains a replicated copy of the ledger, endorses transactions, and hosts smart contracts, known as chaincode, that encode the business logic governing transaction execution.

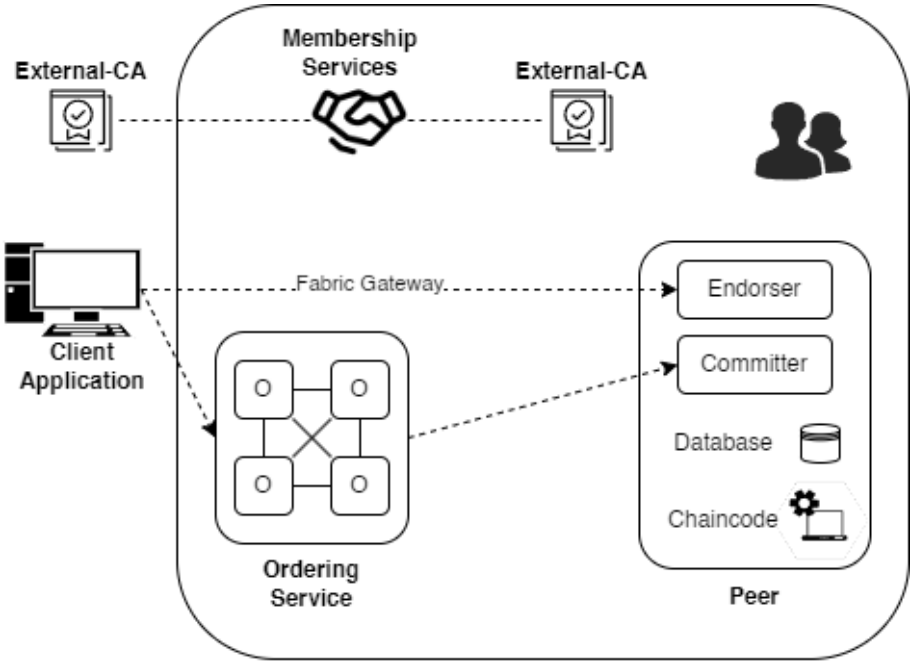


Figure 1. Modular architecture of Hyperledger Fabric, highlighting its key components

The orderer serves as the backbone of transaction management, orchestrating the ordering of transactions and distributing blocks to peers to ensure a consistent and validated transaction sequence across the network. The ledger structure in Hyperledger Fabric is divided into two parts: the world state and the transaction log. The world state represents the current snapshot of all network assets, while the transaction log maintains an immutable record of all

transactions that have shaped the current state. This dual-layer design facilitates efficient asset querying, robust data validation, and comprehensive audit trails for enterprise applications.

2.2. Hyperledger Caliper

Hyperledger Caliper (HLC) is a benchmarking tool designed to evaluate the performance of blockchain networks by generating workload scenarios and collecting metrics that provide performance insights under various conditions [13]. It achieves this by interfacing directly with Fabric networks through the Fabric peer gateway, allowing it to interact with the network like a client application, such as sending transactions to the network and querying the ledger. The architecture of HLC is shown in Figure 2.

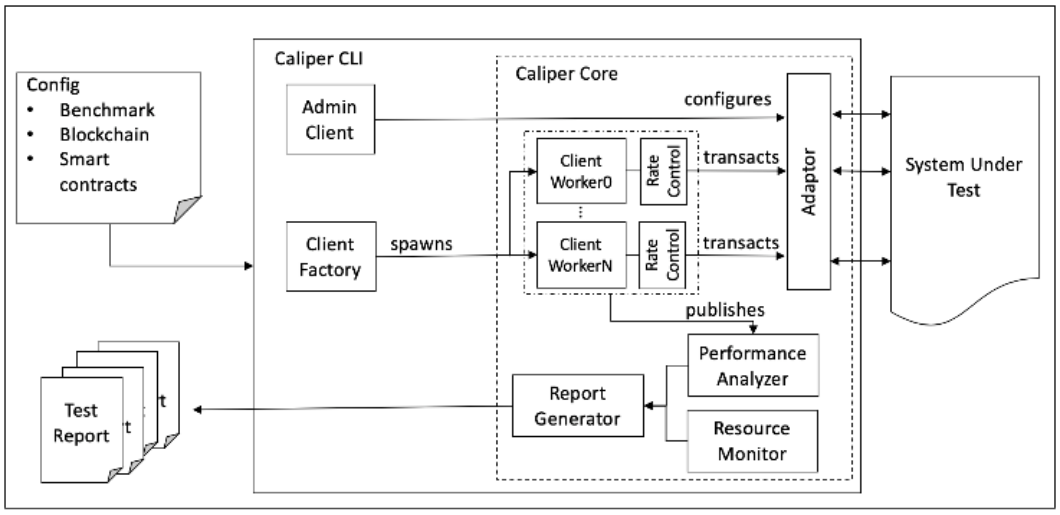


Figure 2. System architecture of Hyperledger Caliper and connection to the System Under Test (SUT) [13]

The benchmarking process begins with configuring the test environment and defining the workload, including a predefined set of transactions and operations to be executed on the blockchain network. Hyperledger Caliper submits transaction proposals to endorsing peers based on the network's defined endorsement policy, using specified network and connection profiles. Peer responses are collected and analyzed to evaluate key performance metrics, including throughput (TPS), which measures the number of transactions processed per second, and latency, which assesses the time required for transaction confirmation. Caliper also monitors memory consumption (usage of peers and orderers), CPU utilization (processing resources consumed), disk read/write (R/W) operations, and network traffic (data exchanged between nodes).

Hyperledger Caliper interacts with Hyperledger Fabric as the system under test (SUT), leveraging the Fabric peer gateway for transaction invocation and response collection. It integrates with tools like Prometheus and Grafana for enhanced monitoring and visualization, offering real-time metrics and dashboard capabilities. Workload definitions and core functionalities are implemented using Node.js and JavaScript, providing a flexible framework for scripting and executing benchmarks.

2.3. Transaction Flow in Hyperledger Fabric

The transaction flow within Hyperledger Fabric is illustrated in Figure 3. The flow follows a structured process to ensure transactions are accurately endorsed, ordered, and committed to the ledger. This process begins with a client application submitting a transaction proposal to

https://doi.org/10.31436/iiumej.v26i3.3610

the endorsing peers (step 1). These peers simulate the transaction by executing the associated chaincode (step 2) and return a proposal response that includes a digital signature and the read/write set of the simulated transaction (step 3). The client application collects and verifies these endorsement responses against the endorsement policy. If the policy is satisfied, the client submits the transaction to the orderer (step 4).

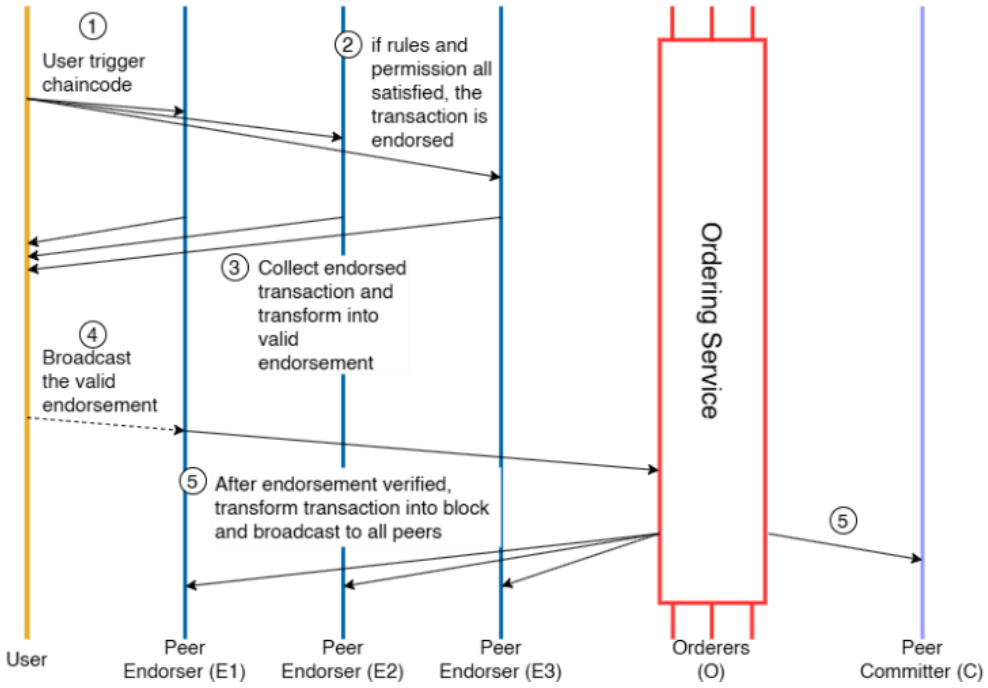


Figure 3. The timing diagram showing the flow of chaincode endorsement and block distribution across different types of nodes in Hyperledger Fabric

The orderer organizes the transaction into a block and disseminates this block to all network peers (step 5). Each peer validates the transaction to ensure it conforms to the endorsement policy and checks for any conflicts in the read/write set. Once validated, the transaction is committed to the ledger, and the world state is updated accordingly.

2.4. Related Works

Research on Hyperledger Fabric's performance and scalability emphasizes its suitability for secure and efficient transaction processing across various applications, including IoT. Comparative studies show that Fabric generally outperforms Ethereum in terms of throughput and latency, though its performance declines under high transaction rates, particularly in IoT healthcare systems [14]. Studies [15], [16] identify key performance factors sensitive to configuration, such as CouchDB utilization and transaction validation speeds, which impact scalability and latency.

Investigations into IoT integration in [17], [18], [19] highlight Fabric's strengths, with implementations in smart cities and enhanced security models improving both throughput and resource efficiency over standard models. In particular, a Raspberry Pi-based access control system demonstrated Fabric's capability in IoT device management. Further studies focusing on data integrity and security underscore Fabric's support for high data security, although scalability challenges remain [20]. Analysis of different node configurations offers developers insights for reducing latency [21].

Empirical research on Fabric's scalability for small and medium enterprises (SMEs) [22] and validated the latency model [23] further contribute to understanding its limitations and potential. While Fabric performs well under moderate workloads, larger-scale deployments require careful optimization to address scalability and performance bottlenecks effectively.

Moreover, several studies and pilot projects have highlighted the application of Hyperledger Fabric in supply chain management and halal certification, particularly in Southeast Asia. Implementations have demonstrated Fabric's ability to ensure product traceability, authenticity, and compliance within complex supply chains, with the added benefit of supporting halal certification processes through transparent and tamper-proof recordkeeping [24], [25]. These use cases further validate Fabric's flexibility in enabling trust, auditability, and process automation in real-world business environments. They also highlight the need for network optimization to manage increased transaction volumes and data complexity.

3. METHODOLOGY

This work implemented a heterogeneous blockchain network comprising nodes with varying hardware architectures and computational resources to evaluate the impact of heterogeneity on performance and scalability. The network consisted of a high-performance laptop, a medium-performance laptop, and two SBCs equipped with ARM Cortex-A72 and Cortex-A76 processors. Table I outlines the hardware specifications and CPU scores, determined using the PassMark Performance Test [26], which evaluates computational capabilities through tasks such as integer and floating-point calculations, data compression, and encryption. These scores provide a standardized measure of processing power, essential for understanding each node's ability to support blockchain operations in Hyperledger Fabric. The methodology began with the design of the blockchain network topology, followed by optimizing orderer settings and developing chaincode for transaction testing. The network was then deployed, and performance measurements were conducted using Hyperledger Caliper.

Table 1. Summary of Hardware Specifications

Device Number	Device Model	Processor	CPU Score	Disc Type	Nodes in the Device	Idle	Sustained Load
1	Acer Aspire 5	Intel Core i7-1355U @1.70 GHz	14820	SSD	2 peers	6.9W	63.4W
2	Acer Aspire R14	Intel Celeron 3205U @ 1.50 GHz	947	HDD	2 peers	6.6W	32.2W
3	Raspberry Pi 4	ARM Cortex-A72 1.8 GHz	692	HDD	2 peers	2.44W	4.84W
4	Raspberry Pi 5	ARM Cortex-A76 2.4 GHz	2151	HDD	3 orderers	4.7W	11.6W

3.1. Blockchain Network Architecture Design

The blockchain network setup consists of four devices, as listed in Table I, each assigned specific roles. Ubuntu was selected as the operating system for all devices to ensure compatibility with Hyperledger Fabric dependencies, including Docker, Docker Compose, Node.js, and the Go programming language. Docker containers were used to host all nodes and services, with Docker Swarm orchestrating the network to create a unified system.

As illustrated in Figure 4, Devices 1, 2, and 3 host the peer nodes, representing high-performance, mid-range, and low-end hardware tiers, respectively. Despite their hardware differences, all peer nodes were uniformly configured, with each device hosting an endorser

https://doi.org/10.31436/iiumej.v26i3.3610

node, a committer node, a dedicated database, and a chaincode container. This standardized configuration across diverse hardware tiers provided a consistent foundation for evaluating the impact of computational resources on transaction throughput, latency, and resource utilization within the network.

3.2. Orderer Configuration Optimization

Before benchmarking individual peer nodes, the configuration of the orderer node in the Hyperledger Fabric network was optimized. This process, a standard best practice in setting up Hyperledger Fabric networks, focused on fine-tuning key parameters to ensure scalability and performance. First, the maximum pending load was adjusted to identify the tipping point of the orderer node without sacrificing throughput. This experiment involved monitoring the number of pending transactions the orderer could handle before performance degradation occurred.

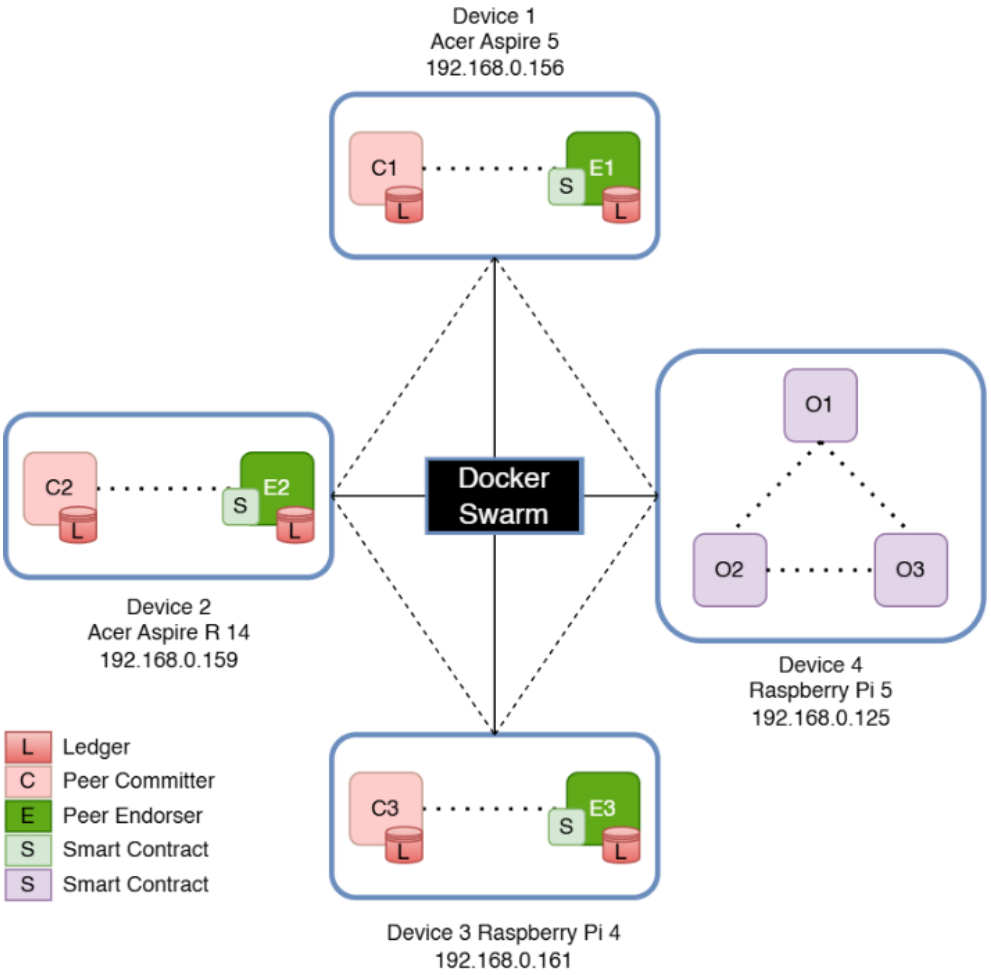


Figure 4. Benchmarked Hyperledger Fabric nodes topology. Devices 1, 2, and 3 hold the same configuration of peer nodes while Device 4 acts as the Orderer organization.

Next, the block size was fine-tuned to manage memory effectively. By adjusting the number of transactions per block, the configuration ensured that each block could accommodate a balanced load without overwhelming the system. Then the batch timeout parameter was fine-tuned by testing various intervals for batching transactions, minimizing delays during low-traffic periods while maintaining efficient block formation under high traffic. These optimized parameters were obtained through iterative experimentation, allowing

for the identification of configurations that optimized the network for a high-performing and consistent baseline.

3.3. Peer Nodes Benchmarking

The benchmarking settings were defined in the caliper workload configuration file (caliper-config.yaml) and network configuration file (network.yaml), specifying details of the blockchain network, chaincodes to invoke, and assigned workload parameters. Three chaincodes were used in this benchmarking exercise to evaluate the network's read, write, and read/write operations capability. These chaincodes include read, tokenize (create a digital asset), and transfer (do a digital transfer), respectively, as illustrated in Figure 5.

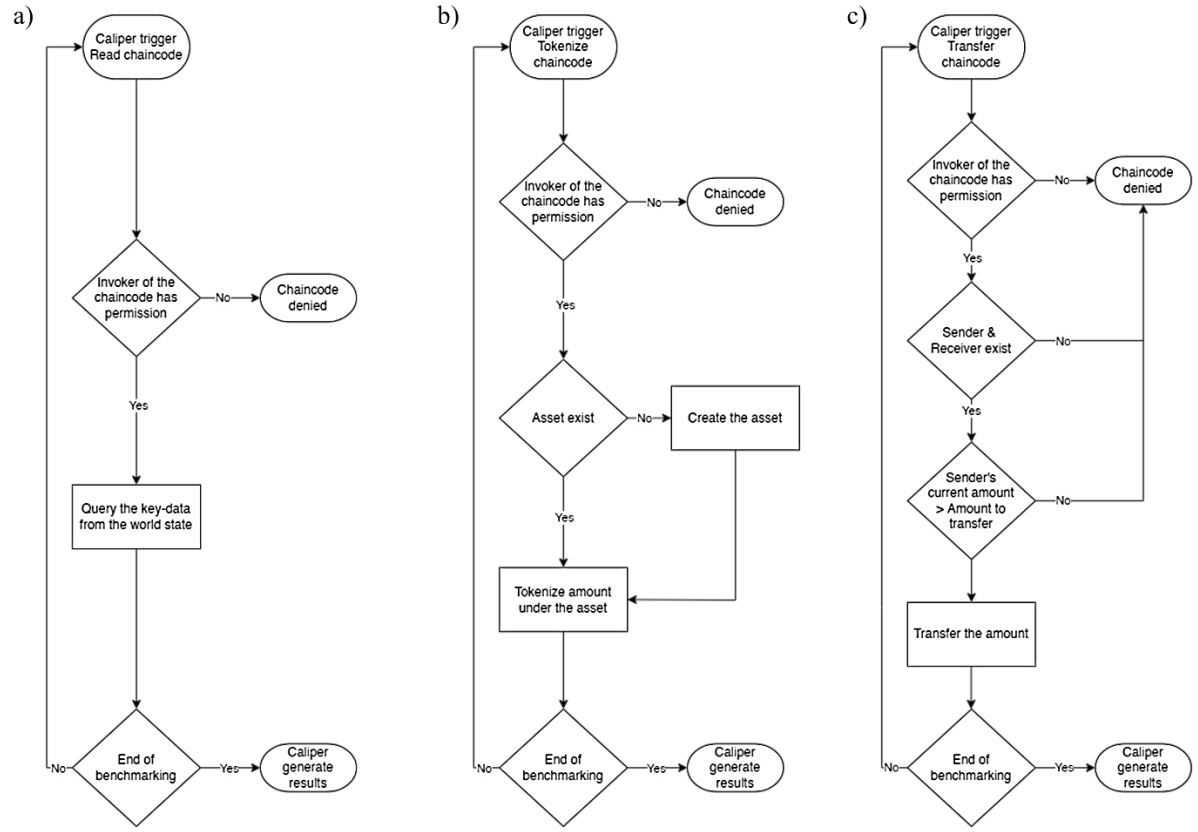


Figure 5. Three Flowcharts of Chaincode Benchmarked. (a) Read chaincode query ledger data from the blockchain world state, (b) Tokenize chaincode, create a new asset if it doesn't exist, or tokenize a new amount to the existing asset, (c) Transfer chaincode check if the sender's amount is higher than the amount to transfer to the sender before doing the transaction

The read chaincode (Figure 5a) retrieves an asset's current state from the state database without generating a new transaction that alters the state database or the blockchain ledger. This chaincode evaluates query performance and provides insights into how it scales with larger datasets and higher concurrency levels.

Figure 5b illustrates the process flow for tokenization, which involves a chaincode that creates a new asset on the blockchain. This consists of writing a new key-value pair to the state database and recording the transaction on the blockchain. The tokenization process tests the network's capacity to handle a large number of asset creation requests in a short time. It analyzes the impact of ledger growth on write-heavy operations and storage performance.

https://doi.org/10.31436/iiumej.v26i3.3610

The third chaincode, transfer, as shown in Figure 5c, updates an existing asset's state to reflect a transfer or change in ownership. This process involves modifying an existing entry in the state database and recording the transaction on the blockchain. It combines read operations (to validate asset ownership or conditions) with write operations (to update the asset state). This chaincode evaluates the network's ability to process complex transactions involving both reads and writes, helping to identify bottlenecks in transaction validation and contention resolution, particularly under high concurrency conditions.

To execute the benchmarking process, the following workload parameters were configured: the Transaction Rate, which defines the rate at which transactions are sent to the blockchain in transactions per second (TPS); the Total Number of Transactions, which specifies the overall volume of transactions to be executed during the test; and the Number of Workers, which represents the total number of concurrent clients generating transactions. These parameters simulate multiple clients interacting with the blockchain and evaluate the network's ability to handle concurrency while maintaining performance.

4. RESULTS AND DISCUSSION

The results section will first discuss the results of the orderer configuration optimization using HLC, where the optimal settings for maximum pending load, block size, and batch timeout can be obtained. These numbers were then used as the orderer parameters. The second part of the results discusses the Caliper measurements for each device running the HLF nodes. The workload settings for each device differ based on its capability to handle the load. The resulting network performance and the resource consumption for each are presented.

4.1. Orderer Configuration Optimization

The tokenize chaincode was used to execute operations for this step, with a fixed transaction load rate of 50 TPS. The simulation involved 100 workers simulating concurrent clients and operating on 10,000 assets. The performance of the orderer node was evaluated using two key metrics, as shown in Figure 6: throughput, the number of successful transactions processed per second (TPS), and success rate, the percentage of submitted transactions successfully processed and committed to the blockchain.

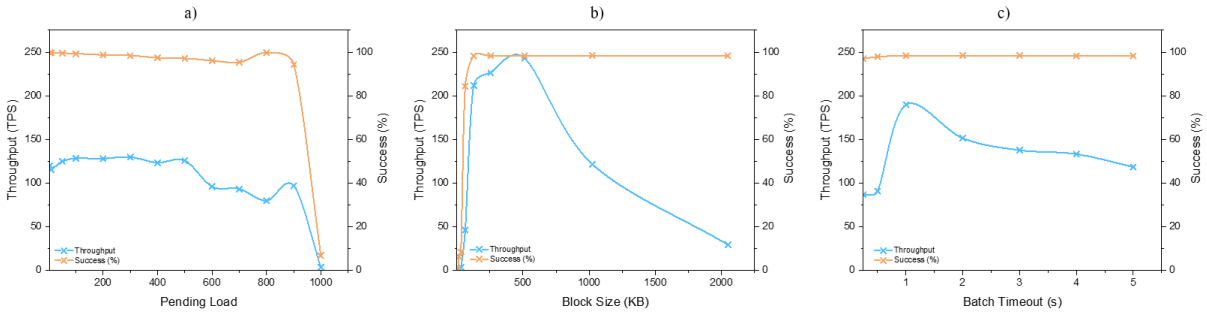


Figure 6. Optimizing configuration for the orderer nodes by modifying and testing: 6a) Pending Load, 6b) Block Size, and 6c) Batch Timeout against Throughput (TPS) and Success (%)

To find the optimum maximum pending load, the simulation was run for 300 seconds. Pending load defines the maximum number of transactions waiting to be grouped into blocks before being sent to peer nodes for validation and commitment. Starting with a pending load of 5, the value was incrementally increased by 100 up to 1,000, while throughput and success

https://doi.org/10.31436/iiumej.v26i3.3610

rate were recorded. As shown in Figure 6a, the orderer could handle a pending load of up to 500, achieving a throughput of approximately 130 TPS while maintaining a success rate above 95%. Beyond this point, latency increases, hence degrading throughput. However, the success rate remained stable until a pending load of 900. At 1,000 pending transactions, the network reached its limit.

Figure 6b shows the effect of block size on network performance. Block size refers to the maximum amount of data in a single block. Starting with an initial block size of 16 KB, the size was incrementally doubled to 2,048 KB. The highest throughput, approximately 243 TPS, was achieved at a block size of 512 KB, while the success rate remained unaffected across all tested sizes.

The effect of batch timeout on performance is shown in Figure 6c. Batch timeout specifies the maximum time the orderer waits before forming a block, even if the block size has not been reached. The simulation tested batch timeouts ranging from 0.25 seconds to 5 seconds. A batch timeout of 1 second gives the highest throughput, with the success rate remaining stable from 1 second to 5 seconds. These tests identified the optimal settings for the orderer node: a pending load of 500, a block size of 512 KB, and a batch timeout of 1 second.

4.2. Peer Nodes Benchmarking Results

The three devices evaluated in this step were Device 1, Device 2, and Device 3, benchmarked across three chaincode operations: read, tokenize, and transfer. Hyperledger Caliper recorded performance metrics, including network performance metrics (throughput and latency) and hardware utilization metrics (CPU usage, memory consumption, and network traffic).

Each device was tested according to its computational capacity. Device 1, a high-performance laptop, handled 10 workers simulating concurrent clients, while Device 2, a mid-range laptop, managed 5 workers. Device 3, a single-board computer (SBC), was assigned just one worker, reflecting its limited capacity and typical use cases where SBCs, such as the Raspberry Pi, are not expected to handle large client loads. The workload settings (TPS) for each device, detailed in Table II, represent the maximum transaction load they could manage as blockchain nodes processing client requests.

Table 2. Transaction Loads In Each Device Per Chaincode

Device	Read	Tokenize	Transfer
1	1500	500	500
2	300	150	150
3	100	50	50

The results are shown in Figure 7, comparing performances across the three devices. As shown in Figure 7a, the high-performance laptop (Device 1) consistently achieves the highest throughput across all chaincode operations, recording 1084.5 TPS for read, 356.8 TPS for tokenize, and 336.8 TPS for transfer. The mid-range laptop (Device 2) delivers lower throughput values, achieving 177.4 TPS, 63.5 TPS, and 19.5 TPS for the same chaincodes. The single-board computer (Device 3), constrained by its limited processing power and storage speed, records peak throughputs of 53.6 TPS, 21.4 TPS, and 18.4 TPS for read, tokenize, and transfer, respectively.

As shown in Figure 7b, Device 1 maintains the lowest latency, with a maximum of approximately 3 seconds for the tokenize chaincode. Devices 2 and 3 exhibit higher latencies,

https://doi.org/10.31436/iiumej.v26i3.3610

although both remain below 17 seconds, even for more complex operations such as tokenization and transfer. While Device 2, the mid-range laptop, experiences higher latency than Device 3, the SBC, it is essential to note that Device 2 handles a significantly larger workload and a greater number of workers than Device 3. In comparison, public blockchains such as Bitcoin and Ethereum typically exhibit much higher latencies, with Bitcoin averaging around 10 minutes per transaction and Ethereum approximately 15 seconds per block confirmation.

To better contextualize CPU performance, CPU utilization was compared across devices at their respective optimal workloads, illustrated in Figure 7c. For the transfer chaincode, Device 1 maintained ~6% CPU usage at 336.8 TPS, Device 2 reached ~25% at only 63.5 TPS, and Device 3 recorded ~23% at 18.4 TPS. While the CPU usage of Devices 2 and 3 is relatively close, their throughput differs significantly, and the effective processing throughput per percentage of CPU usage is far higher on Device 1 due to its faster clock speed, larger cache, and more efficient architecture.

The primary contributors to CPU load were cryptographic signing and verification (ECDSA), chaincode execution involving read/write operations, and ordering service block processing. Two operational profiles emerged from the measurements: (1) blockchain services alone consumed 20–25% of available CPU, and (2) when operating system processes, Docker overhead, and Caliper benchmarking were included, CPU utilization approached maximum capacity. This confirms that CPU contention is a limiting factor during high-TPS workloads.

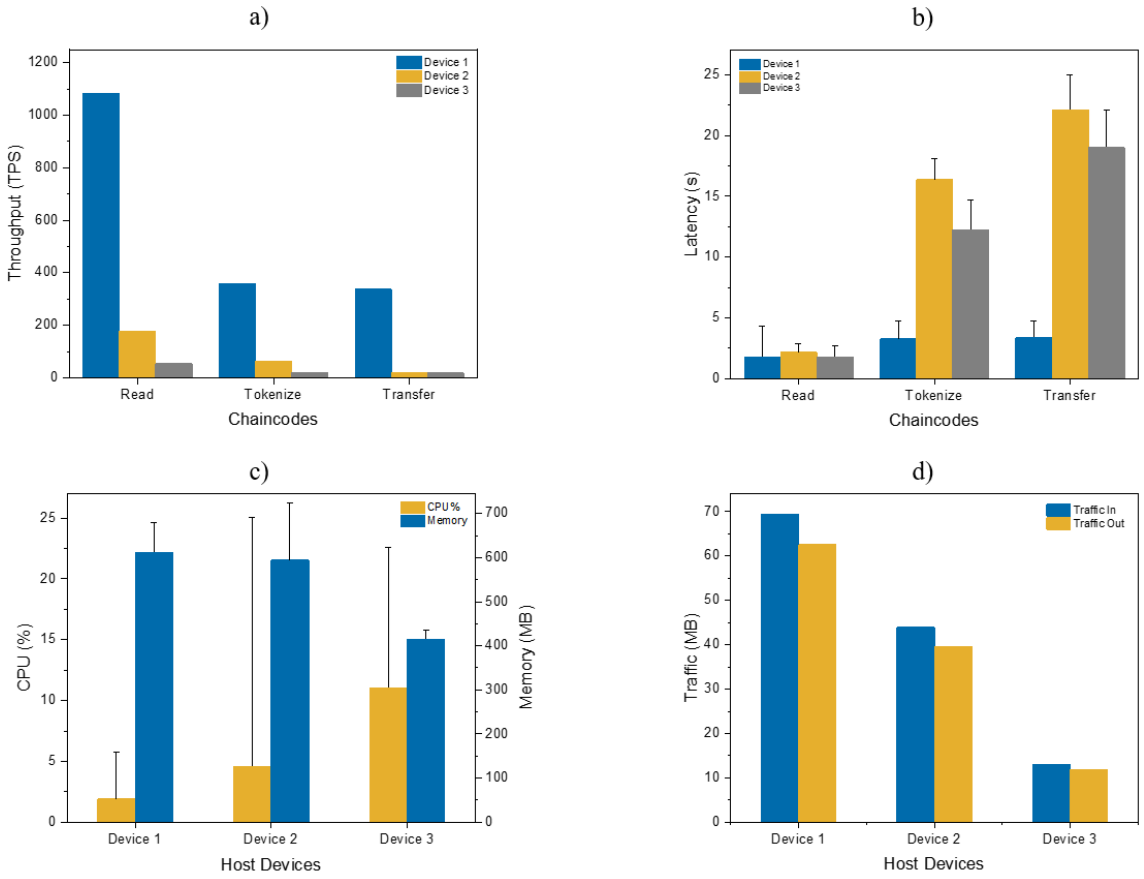


Figure 7. Performance Benchmarking Results: (a) chaincode performance throughput (TPS), (b) chaincode latency (s), (c) hardware usage during benchmarking, and (d) traffic inflow and outflow

https://doi.org/10.31436/iiumej.v26i3.3610

Beyond CPU utilization, network and storage performance played a significant role. All devices were connected via Wi-Fi to a single router, introducing bandwidth sharing and occasional latency spikes. Device 1's SSD allowed faster ledger writes compared to HDD-equipped Device 2, contributing to its lower latency. Device 3's limited network throughput (~4.36 MB/s inflow), combined with slower storage, increased failure rates under heavy load. These factors have direct implications for IoT deployments — in bandwidth-constrained or high-latency environments, transaction confirmation delays could disrupt real-time IoT applications such as sensor event logging or automated control systems.

Transaction payload size and chaincode complexity also impact performance. While the current tests used small payloads, IoT deployments often involve larger datasets or batch transactions. Larger payloads increase serialization, transmission, and cryptographic processing time. Complex chaincodes (e.g., multi-asset transfer with conditional checks) consume more CPU cycles per transaction and extend endorsement times under high concurrency, especially on low-power devices.

Regarding network capability, Figure 7d illustrates the network traffic across devices, measured in megabytes per second (MB/s) over the 3-second benchmark duration. Device 1 exhibits the highest network throughput, with an inflow rate of approximately 23.18 MB/s and an outflow rate of 20.86 MB/s. Device 2 exhibits moderate network usage, with an inflow rate of 14.64 MB/s and an outflow rate of 13.18 MB/s. In contrast, Device 3 experiences significantly lower network traffic, with an inflow rate of 4.36 MB/s and an outflow rate of 3.93 MB/s. This reduced throughput for Device 3 is attributed to increased failure rates when handling high transaction volumes within the short benchmarking period. These results underscore the varying capabilities of each device, suggesting that high-performance nodes are essential for applications with intensive workloads. In contrast, lower-powered devices can still contribute as supportive nodes, effectively balancing the network by handling moderate transactions. The correlation between network throughput and overall TPS suggests that sufficient bandwidth is critical for high-performance blockchain nodes in IoT contexts.

Power efficiency is another critical factor for IoT-blockchain deployments, particularly in regions like Malaysia, where environmental sustainability initiatives aim to reduce energy footprints. Table 1 presents the idle and sustained load power consumption for all devices. When normalized against peak transaction throughput, high-performance devices like Device 1 achieve more TPS per watt under heavy load (≈17.1 TPS/W for the tokenize chaincode) than lower-power devices like Device 3 (≈4.4 TPS/W). This highlights a trade-off: while SBCs consume far less energy overall, their reduced processing capacity lowers TPS efficiency.

From a sustainability perspective, deployment strategies in Malaysia could balance these trade-offs by assigning high-throughput workloads to energy-efficient high-performance nodes and reserving low-power devices for lighter tasks. This approach supports blockchain adoption in green initiatives such as carbon offset tracking, renewable energy certificates, and transparent supply chain management, aligning with Malaysia's environmental goals while minimizing unnecessary energy usage.

4.3. Optimization Considerations

Optimizing heterogeneous networks for IoT requires role allocation that matches device capability. SBCs like Raspberry Pi can serve as commit-only peers or lightweight endorsers to reduce cryptographic load. Switching from CouchDB to LevelDB reduces query latency, and lightweight Docker images with fewer background services lower CPU and memory usage. Data preprocessing at the edge before blockchain submission can also minimize transaction

frequency and payload size. For compute-intensive environments, hardware accelerators such as GPUs or FPGAs could offload signature verification and chaincode execution, boosting throughput without replacing existing hardware.

5. CONCLUSION AND FUTURE WORK

This research explored the performance characteristics of a multi-device Hyperledger Fabric network, deploying a diverse mix of hardware to evaluate blockchain and hardware performance under different configurations. Key insights from the benchmarking tests indicate that while the network demonstrates effective transaction handling across devices, factors such as device processing capabilities and CPU utilization play significant roles in determining overall network efficiency.

One notable observation involves the CPU utilization, which appears modest at approximately 20% - 25% when considering only the blockchain services. However, this measurement does not account for CPU usage by the operating system or the Hyperledger Caliper benchmarking tools. When these additional processes are included, the CPU approaches maximum capacity, indicating that CPU limitations are the primary bottleneck affecting network performance. This contention for CPU resources restricts the system's ability to handle higher transaction volumes efficiently. Optimizing processing capabilities, such as upgrading to more powerful CPUs or enhancing the efficiency of blockchain services, could alleviate these limitations and increase throughput capacity.

Optimization strategies for low-cost devices are crucial to improve further accessibility and adoption of Hyperledger Fabric among SMEs and rural communities. Assigning less resource-intensive node roles, such as peer-only nodes, to low-spec devices, while reserving more demanding services for powerful hardware, can enhance efficiency without increasing cost. Streamlining chaincode logic minimizes computational complexity and reduces the need for large payloads, thereby alleviating the resource burden. Additionally, preprocessing and aggregating data at the edge before submission helps distribute compute load and lessen the number of blockchain transactions required, particularly in resource-constrained environments. Lightweight container configurations and simple consensus mechanisms further contribute to overall system efficiency. These approaches collectively make blockchain deployment more achievable for organizations with limited infrastructure.

Future work could assess hardware configurations and optimization strategies to mitigate CPU bottlenecks and enhance performance. Expanding testing to include more complex or resource-intensive chaincodes may provide insights into how processing demands interact with hardware setups under various conditions. This study provides a foundational understanding that will be valuable for designing Hyperledger Fabric networks on diverse hardware configurations. The planned optimizations are expected to yield further efficiency gains in future iterations.

REFERENCES

[1] Laroui M, Nour B, Moungla H, Cherif MA, Afifi H, Guizani M. (2021). Edge and fog computing for IoT: A survey on current research activities & future directions. Computer Communications, 180, 210-231.

[2] Maraveas C, Piromalis D, Arvanitis KG, Bartzanas T, Loukatos D. (2022). Applications of IoT for optimized greenhouse environment and resources management. Computers and Electronics in Agriculture, 198, 106993.

[3] Kumar NM, Chand AA, Malvoni M, Prasad KA, Mamun KA, Islam FR, Chopra SS. (2020). Distributed energy resources and the application of AI, IoT, and blockchain in smart grids. Energies, 13(21), 5739.

[4] Mathur S, Kalla A, Gür G, Bohra MK, Liyanage M. (2023). A survey on role of blockchain for IoT: Applications and technical aspects. Computer Networks, 227, 109726.

[5] Makhdoom I, Abolhasan M, Abbas H, Ni W. (2019). Blockchain's adoption in IoT: The challenges, and a way forward. Journal of Network and Computer Applications, 125, 251-279.

[6] Ashley MJ, Johnson MS. (2018). Establishing a secure, transparent, and autonomous blockchain of custody for renewable energy credits and carbon credits. IEEE Engineering Management Review, 46(4), 100-102.

[7] Anand P, Singh Y, Selwal A, Alazab M, Tanwar S, Kumar N. (2020). IoT vulnerability assessment for sustainable computing: Threats, current solutions, and open challenges. IEEE Access, 8, 168825-168853.

[8] Siwakoti YR, Bhurtel M, Rawat DB, Oest A, Johnson RC. (2023). Advances in IoT security: Vulnerabilities, enabled criminal services, attacks, and countermeasures. IEEE Internet of Things Journal, 10(13), 11224-11239.

[9] Sharma PK, Kumar N, Park JH. (2020). Blockchain technology toward green IoT: Opportunities and challenges. IEEE Network, 34(4), 263-269.

[10] Honar PH, Rashid MA, Alam F, Demidenko S. (2022). Experimental performance analysis of a scalable distributed hyperledger fabric for a large-scale IoT testbed. Sensors, 22(13), 4868.

[11] Kaushal RK, Kumar N. (2024, March). Exploring hyperledger caliper benchmarking tool to measure the performance of blockchain based solutions. In 2024 11th international conference on reliability, infocom technologies and optimization (trends and future directions)(ICRITO) (pp. 1-6). IEEE.

[12] Introduction - Hyperledger Fabric Docs main documentation. [https://hyperledger-fabric.readthedocs.io/en/release-2.5/whatis.html]

[13] Introduction - Hyperledger Caliper. [https://hyperledger-caliper.github.io/caliper/0.6.0/]

[14] Ucbas Y, Eleyan A, Hammoudeh M, Alohaly M. (2023). Performance and scalability analysis of ethereum and hyperledger fabric. IEEE Access, 11, 67156-67167.

[15] Guggenberger T, Sedlmeir J, Fridgen G, Luckow A. (2022). An in-depth investigation of the performance characteristics of Hyperledger Fabric. Computers & Industrial Engineering, 173, 108716.

[16] Wang C, Chu X. (2020, November). Performance characterization and bottleneck analysis of hyperledger fabric. In 2020 IEEE 40th International Conference on Distributed Computing Systems (ICDCS) (pp. 1281-1286). IEEE.

[17] Alexandridis A, Al-Sumaidaee G, Zilic Z, Jeon G, Wang J. (2023). An iot ecosystem platform for the evaluation of blockchain feasibility. IEEE Internet of Things Journal, 10(24), 21515-21527.

[18] Honar PH, Rashid M, Alam F, Demidenko S. (2021). Hyperledger fabric blockchain for securing the edge internet of things. Sensors, 21(2), 359.

[19] Iftekhar A, Cui X, Tao Q, Zheng C. (2021). Hyperledger fabric access control system for internet of things layer in blockchain-based applications. Entropy, 23(8), 1054.

[20] Eghmazi A, Ataei M, Landry RJ, Chevrette G. (2024). Enhancing IoT data security: Using the blockchain to boost data integrity and privacy. IoT, 5(1), 20-34.

[21] Xu X, Sun G, Luo L, Cao H, Yu H, Vasilakos AV. (2021). Latency performance modeling and analysis for hyperledger fabric blockchain network. Information Processing & Management, 58(1), 102436.

[22] Khan D, Jung LT, Hashmani MA, Cheong MK. (2022). Empirical performance analysis of hyperledger LTS for small and medium enterprises. Sensors, 22(3), 915.

[23] Foschini L, Gavagna A, Martuscelli G, Montanari R. (2020, June). Hyperledger fabric blockchain: Chaincode performance analysis. In ICC 2020-2020 IEEE International Conference on Communications (ICC) (pp. 1-6). IEEE.

[24] Alamsyah A, Hakim N, Hendayani R. (2022). Blockchain-based traceability system to support the Indonesian halal supply chain ecosystem. Economies, 10(6), 134.

[25] Song S, Lu J, Zhao H, Wang W, Shi C, Luo R. (2022, November). Traceability of Product Supply Chain Based on Hyperledger Fabric. In Proceedings of the 4th International Conference on Advanced Information Science and System (pp. 1-6).

[26] PassMark CPU Benchmarks - CPU Test Information. [https://www.cpubenchmark.net/cpu_test_info.html]

https://doi.org/10.31436/iiumej.v26i3.3612

CHALLENGES AND STRATEGIES FOR IMPLEMENTATION OF LEAD-ACID BATTERY HEALTH MONITORING AND PREDICTION IN OFF-GRID SOLAR PANEL SYSTEMS

YOGA PRASTIYA WIBAWA[1*], I PUTU ANANTA YOGISWARA[1], BONDAN SUWANDI[2], ANDI KURNIANTO[1], BUDI SULISTYA[1], ABHIMATA AR-RASYIID[1], NASHRULLAH TAUFIK[1], ARIEF RUFIYANTO[1], I MADE ASTAWA[1]

[1]*Research Centre of Electronics, National Research & Innovation Agency, Indonesia*
[2]*Directorate of Laboratory Management, Research Facilities, Science and Technology Areas, National Research & Innovation Agency, Indonesia*

**Corresponding author: yoga002@brin.go.id, yoga.prastiya@gmail.com*

(Received: 17 February 2025; Accepted: 9 June 2025; Published online: 9 September 2025)

ABSTRACT: This study presents an IoT-based real-time battery health monitoring system that integrates an Equivalent Circuit Model (ECM) with a linear regression approach to estimate internal resistance (IR) and open-circuit voltage (VOC). Unlike conventional electrochemical models or machine learning-based solutions, this method offers a computationally efficient yet accurate approach to predicting battery degradation. The findings demonstrate that increasing IR correlates strongly with declining battery health, reinforcing the necessity of continuous monitoring for predictive maintenance. Additionally, this research highlights the impact of environmental conditions, particularly temperature fluctuations, on battery efficiency, challenging the assumption that heavy rainfall significantly lowers ambient temperatures. Furthermore, this study addresses real-world implementation challenges, including data loss from network disruptions, by proposing a fault-tolerant IoT architecture. These advancements contribute to smart energy storage and provide a scalable solution for remote and off-grid applications, particularly in smart farming.

ABSTRAK: Kajian ini memperkenalkan sistem pemantauan kesihatan bateri masa nyata berasaskan IoT yang terkini, mengintegrasikan Model Litar Ekuivalen (ECM) dengan pendekatan regresi linear untuk menganggarkan rintangan dalaman (IR) dan voltan litar terbuka (VOC). Berbeza dengan model elektrokimia konvensional atau kaedah pembelajaran mesin yang memerlukan sumber pengiraan tinggi, pendekatan ini menawarkan keseimbangan antara ketepatan dan kecekapan pemprosesan dalam meramalkan degradasi bateri. Hasil kajian menunjukkan bahawa peningkatan IR berkait rapat dengan kemerosotan kesihatan bateri, menegaskan kepentingan pemantauan berterusan untuk penyelenggaraan prediktif. Selain itu, kajian ini meneliti kesan persekitaran terhadap prestasi bateri, terutamanya variasi suhu, dan mencabar tanggapan bahawa hujan lebat secara signifikan menurunkan suhu ambien. Kajian ini juga menangani cabaran pelaksanaan di dunia nyata, termasuk kehilangan data akibat gangguan rangkaian, dengan mencadangkan seni bina IoT yang tahan gangguan. Kemajuan ini menyumbang kepada pembangunan sistem storan tenaga pintar dan menawarkan penyelesaian yang berskala untuk aplikasi luar grid, terutamanya dalam bidang pertanian pintar.

KEYWORDS: *Lead-acid battery, IoT monitoring, internal resistance, open-circuit voltage, state of health(SOH).*

https://doi.org/10.31436/iiumej.v26i3.3612

1. INTRODUCTION

Rapid advancements in the development of electric vehicles or renewable energy smart power systems, such as smart farming, will require efficient and reliable battery systems to progress. In the field of smart farming, where autonomous systems, sensors, and other IoT devices are indispensable, compromising battery health can reduce availability and pose safety risks. However, the complex nature of battery degradation, influenced by internal mechanisms, manufacturing variations, and the challenging environmental conditions of farming operations, complicates accurate lifetime prediction and health management [1].

Lithium-based batteries have emerged as the most preferred option in many applications due to their lightweight, high energy density, and relatively long lifespan. Compared to lithium-based counterparts, lead-acid batteries generally have a lower energy density and shorter lifespan [2]. Nevertheless, they are still deployed in various applications due to their cost-effectiveness, proven reliability, and robustness. That often makes them an economical option, especially for farming fields or stationary power system applications, where the battery does not need to be removed more frequently.

The lead-acid battery degradation mechanisms, including sulfation, grid corrosion, and electrolyte stratification, have been extensively investigated. There are developed strategies to mitigate these issues and improve these batteries' overall performance and durability [3]. However, predicting the lifespan and managing the health of lead-acid batteries remain complex tasks. This complexity arises from a combination of factors, including intricate internal degradation processes, variations introduced during manufacturing, and the often harsh and unpredictable environmental conditions encountered in farming operations [4]. These challenges are key to ongoing research and innovation in improving battery management systems and predictive modeling. The battery monitors several parameters, such as open circuit voltage and internal resistance [5], essential for developing advanced battery management systems [6].

Implementing a combination of model-based and experimental techniques may have made significant progress in this area, leading to a more reliable and efficient battery-powered system. As batteries age, the model parameters must be updated to reflect their performance accurately. This process relies on the limited sensor data in battery health prediction systems. Optimal sensor deployment is key to estimating battery degradation precisely and effectively over time. High rainfall can worsen this challenge by increasing cloud cover, which reduces solar energy generation and forces the battery to discharge more frequently to meet load demands. Consequently, the increased depth of discharge and cycling accelerates battery degradation.

This study contributes to renewable energy storage and battery management by providing real-world insights into battery degradation patterns and practical challenges. Those challenges and strategies for implementing lead-acid battery health monitoring are delivered especially in remote IoT monitoring. The findings will pave the way for further research on adaptive battery management strategies and provide data parameters for predictive maintenance. In the future, this will result in better cost-effectiveness of battery monitoring and accuracy in predicting battery performance degradation.

2. RELATED WORKS

One approach involves modeling the battery using an electrochemical model and real-time data simulation using fuzzy logic, where the model parameters are estimated online [7].

https://doi.org/10.31436/iiumej.v26i3.3612

Machine learning methods can provide insights into battery degradation [8], but implementation can be challenging due to the computational complexity. Internal resistance measurement can also increase the accuracy of State of Charge (SOC) estimation and car mileage [9].

A direct resistance estimation (DRE) technique offers a low-computation alternative that does not require training data [10]. The DRE method estimates ohmic resistance only during sharp current pulses, enabling efficient degraded cell identification within a large battery pack. Results show that DRE maintains high accuracy while significantly reducing computational complexity and advancing diagnostic techniques for battery health monitoring. An alternative approach is to model the battery using an equivalent circuit model and estimate the circuit parameters online. The internal resistance of a battery has proven to increase with aging and can serve as an indicator of the battery's health.

In an Equivalent Circuit Model (ECM), it is essential to identify thermodynamic parameters, such as impedance, and kinetic parameters, including Open Circuit Voltage (VOC). Impedance is particularly significant for State of Health (SOH) diagnosis, as it provides insights into the internal condition of the battery. The ECM offers notable advantages due to its relative simplicity compared to more complex physical models, facilitating more straightforward implementation in simulation software [11]. Moreover, the flexibility of the ECM allows it to be adapted for various types of batteries and applications, including electric vehicles and stationary energy storage systems.

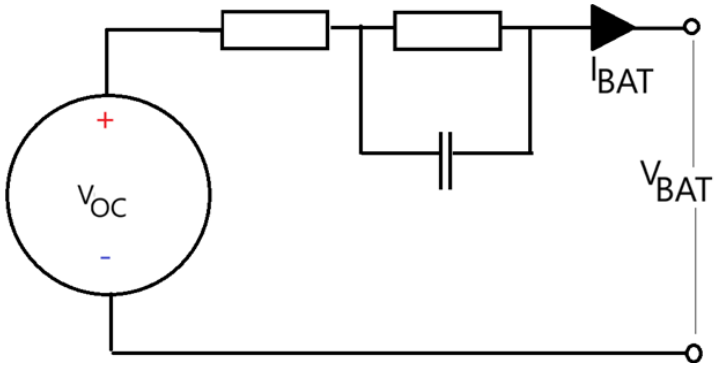


Figure 1. Equivalent circuit model.

The ECM is a widely used approach for representing the behavior of batteries, as shown in Figure 1 [12]. Using this model can help understand the voltage-current (V-I) relationship and the SOH of the battery. By using electrical circuit elements such as resistors, capacitors, and voltage sources, the ECM can simulate the dynamics of the voltage-current characteristic of a battery cell. Characteristics and components of ECM:

- *Resistors and Capacitors*: Resistors cause energy loss due to their resistance, while capacitors can represent the dynamic behavior of the battery during current changes. The impedance analysis may reveal information about the battery's State of Charge (SOC), aging, and potential failure mechanisms, making it a valuable tool for battery management.
- *Voltage Source*: This represents the energy stored by the battery based on its SOC and SOH.

The internal resistance and temperature relationship in commercial batteries is a critical factor influencing their performance. Some research shows that a 4 Ah lead-acid battery with a 70% SOH at room temperature (20 – 35°C) was found to have an internal resistance range

https://doi.org/10.31436/iiumej.v26i3.3612

between 150 and 200 mΩ [13]. The internal resistance tends to increase due to reduced ion mobility within the electrolyte at lower temperatures, which resists electrical current flow. This condition leads to lowered battery performance, decreased efficiency, and reduced power output. The increased internal resistance at lower temperatures is a well-documented challenge in battery performance. Studies have shown that temperature variations significantly impact the electrochemical processes within batteries, particularly in terms of ion transport and reaction kinetics [14]. For instance, diminished ion mobility at lower temperatures results in higher resistance, which reduces the battery's ability to deliver power and accelerates degradation over time. This highlights the importance of maintaining optimal operating temperatures to ensure consistent performance and longevity of batteries.

3. METHODOLOGY

The methodology consists of several key phases.

3.1. Data Acquisition

To develop a data-driven model for predicting battery capacity fade and lifespan, incorporating various environmental factors, internal mechanisms, and manufacturing variations specific to battery types used in smart farming. The data acquisition site is located in a mushroom farming area in Cikuda Village, Parung Panjang, near the South Tangerang region, Indonesia, at an elevation of 866 meters above sea level. Measurement data was recorded from November to December 2024. The monitoring station is located in our laboratory, Gedung Teknologi 3, KST B.J. Habibie. The distance between the data acquisition site and the monitoring station is approximately 8 km, as shown in Figure 2.

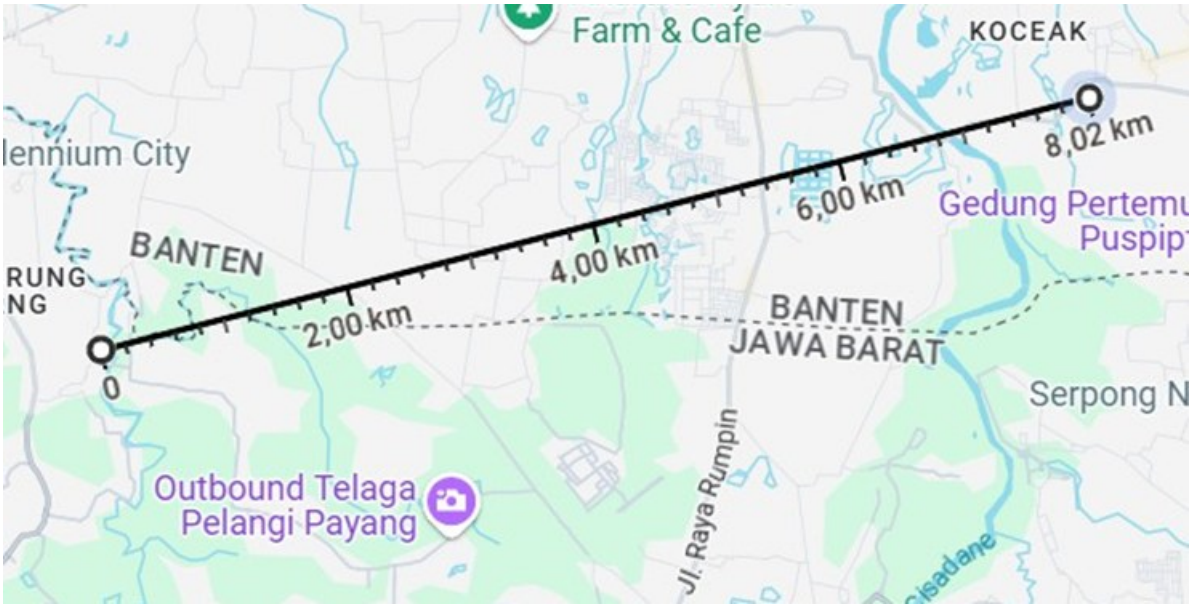


Figure 2. The distance of the measurement location to the monitoring station based on Maps Data © 2025 Google.

3.1.1. Instrumentation Setup

The system utilizes a battery bank connected to a Maximum Power Point Tracking (MPPT) board, serving as a solar charge controller. The battery bank consists of two 12 V, 100 Ah Valve Regulated Lead Acid (VRLA) batteries connected in series. This maintenance-free battery has a rated value of 1.80 V/ cell at 25 °C and an operating temperature range of -15 to

https://doi.org/10.31436/iiumej.v26i3.3612

50 °C. The MPPT board plays a crucial role in managing the charging and discharging of the batteries while simultaneously measuring key electrical parameters, including voltage and current. Measurement and data recording are conducted in real time using IoT technology, as shown in Figure 3.

3.1.2. Data flow

The voltage and current data are transmitted via a serial connection from the MPTT board to a single-board computer (e.g., Raspberry Pi). It processes the information and forwards it to an MQTT broker via cellular telecommunication service over TCP (Internet). A Node.js-based data collector retrieves the data from the MQTT broker and stores it in a PostgreSQL database with a Timescale extension, optimized for time-series data. The entire system is containerized using Docker, ensuring scalability and consistent deployment. This architecture enables efficient and reliable real-time remote measurement.

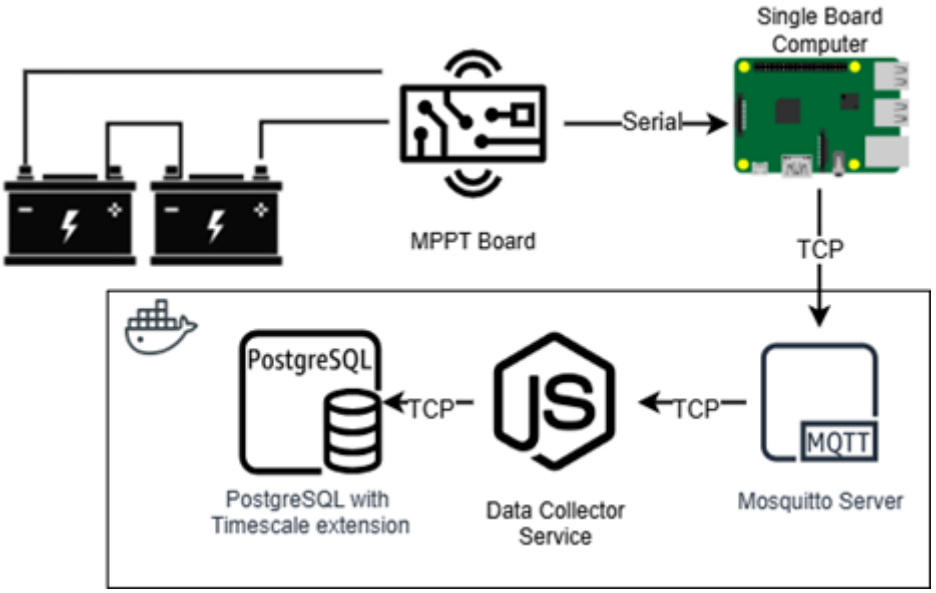


Figure 3. Data acquisition flow chart.

3.2. Data Analysis

To establish a reliable methodology for estimating Open Circuit Voltage (VOC) and internal resistance (IR) under farming environmental conditions using battery voltage and current data.

3.2.1. Data Processing

Data processing is done by analyzing the relationship between the measured voltage and current, as shown in Figure 4, based on direct resistance estimation. Linear regression is an appropriate method for analyzing the relationship between measured current (I_{bat}) and voltage (V_{bat}) in battery systems due to its simplicity, effectiveness, and strong statistical foundation. The voltage-current relationship for a battery under charged-discharge conditions is modeled as a linear Eq. (1).

$$V_{bat} = V_{oc} \pm I_{bat} \cdot R_{int} \quad (1)$$

where V_{bat} notation has meaning, plus when the battery is in a charged condition, and minus when the battery is in a discharged condition. R_{int} is the internal resistance of the battery.

https://doi.org/10.31436/iiumej.v26i3.3612

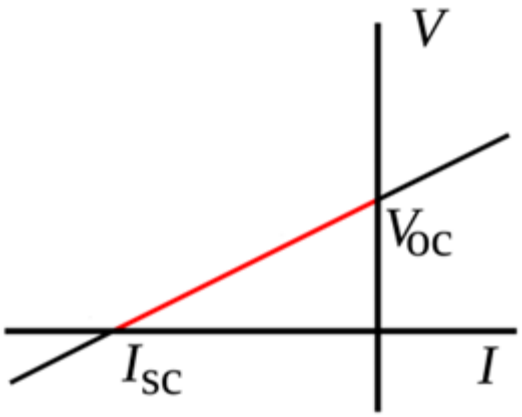


Figure 4. IV Curve relationship.

3.2.2. Data Interpretation

The data interpretation method has a straightforward method that provides easily interpretable results:

- The slope corresponds to R_{int} allowing for direct calculation of the internal resistance (IR).
- The y-intercept corresponds to V_{oc}, representing the open circuit voltage (VOC).
- The x-intercept corresponds to I_{sc}, representing the short circuit current (ISC).

The computational simplicity of linear regression ensures that it can be implemented efficiently, even with large datasets. This efficiency is particularly beneficial in scenarios where extensive cycling data must be analyzed with limited computational resources.

4. RESULTS AND DISCUSSION

Figure 5 provides a clear visual representation of environmental temperature patterns recorded across various dates and times. The vertical axis represents the progression of days, while the horizontal axis captures the time of day. A vivid color gradient, ranging from blue to red, illustrates the temperature spectrum, with cooler readings around 26°C (blue) at midnight transitioning to hotter readings near 36°C (red) at midday. This combination of axes and color coding offers an intuitive overview of temperature fluctuations over time.

Despite the overall consistency of the data, several interruptions were observed. Specifically, data gaps were recorded on November 8, 14, and 15, as well as on December 4, 8, 9, 14, 15, and 16, as shown in Figure 6. In this figure, the availability and continuity of data are represented in the ratio of the number of data points received daily over the maximum number of data points during measurement periods. Furthermore, a complete absence of daily data was noted between November 25 and 28, and again on December 15. These disruptions primarily stem from server downtimes and data transmission issues, common obstacles in large-scale data acquisition systems. While such interruptions may pose temporary challenges, once server functionality is restored and communication channels are re-established, data flow resumes to rerun automatically.

Furthermore, this situation also tests the reliability of our monitoring system, which utilizes public telecommunications and other third-party services. The challenges during this period highlight the importance of robust unmanned and remote data collection infrastructure

https://doi.org/10.31436/iiumej.v26i3.3612

and the need for contingency plans to mitigate the impact of such disruption. This is practical as it eliminates the need to leave highly valuable data recorder equipment unsupervised by on-location personnel and exposed to the weather.

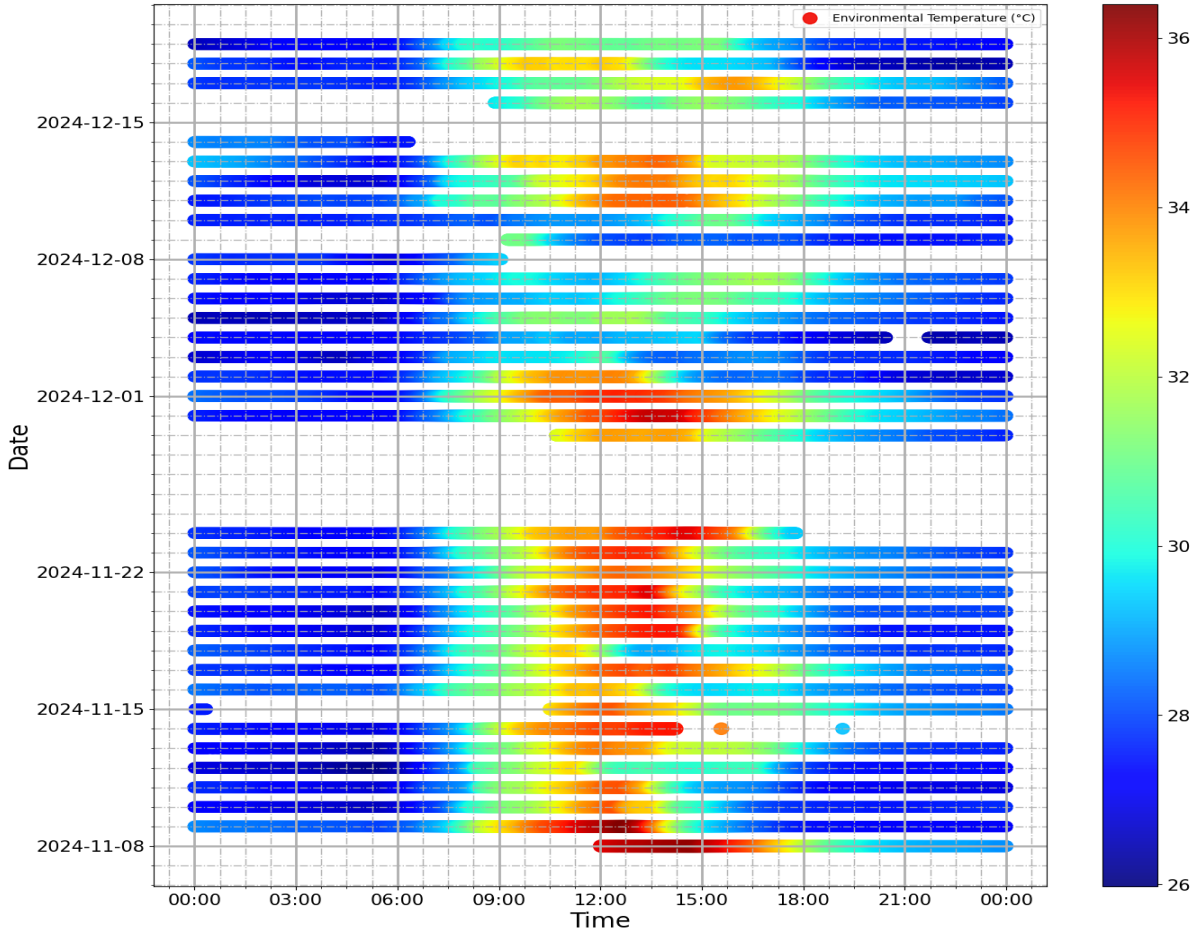


Figure 5. Daily battery environmental temperature record during data acquisition.

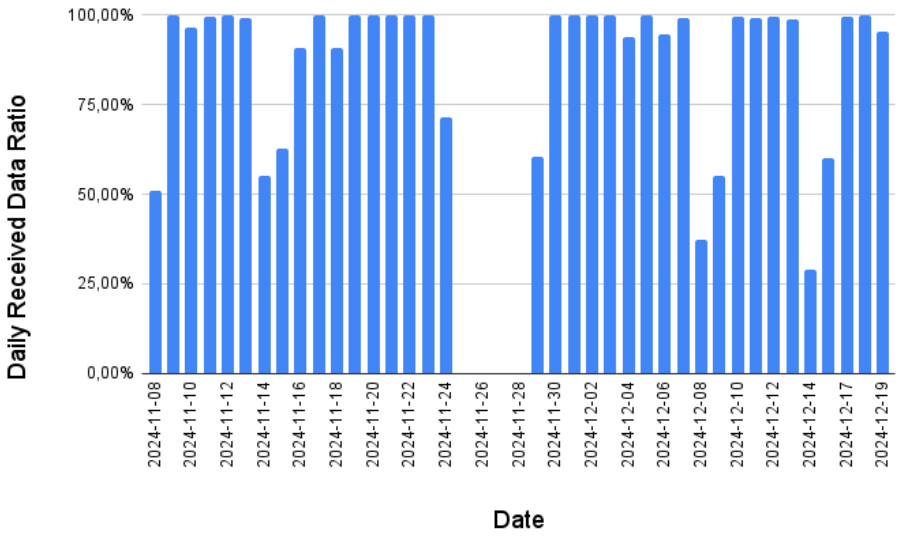


Figure 6. Daily number of data recording.

https://doi.org/10.31436/iiumej.v26i3.3612

Figure 7 illustrates the cycle performance of a battery system over time, presented in 7 days of observation in sub-figures, characterized by its behavior in terms of voltage (plotted on the vertical axis) and current (plotted on the horizontal axis). Each line represents a daily cycle, with different colors and markers corresponding to specific dates. The plot captures the charging and discharging patterns of the battery, providing a comprehensive view of its operation across multiple days. A notable clustering of data points is observed around 0 amperes and a voltage range of 26 to 28.5 volts. This concentration suggests that the battery often remains in a steady-state condition, likely during rest periods or when fully charged.

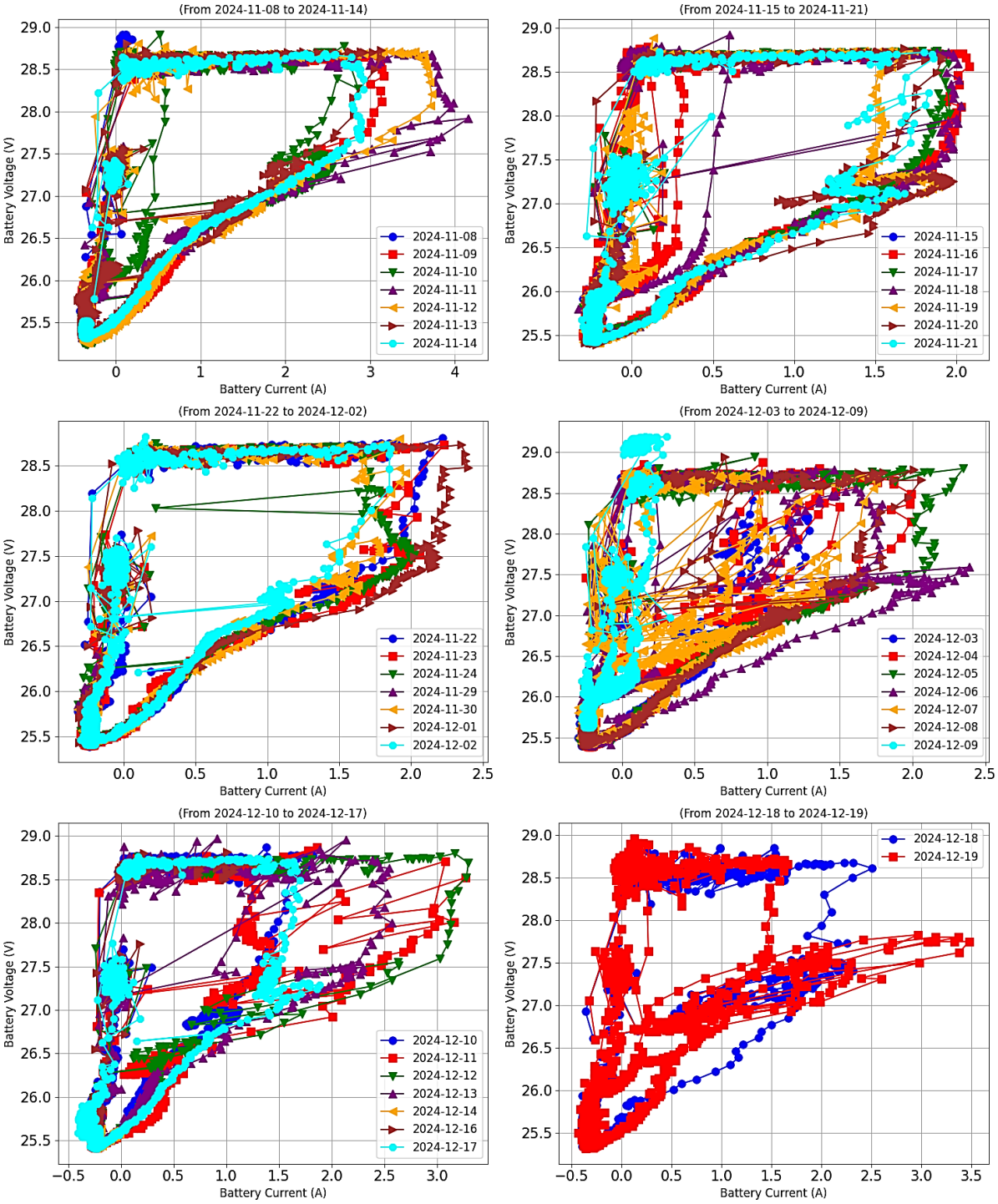


Figure 7. Daily Battery Cycle.

https://doi.org/10.31436/iiumej.v26i3.3612

Figure 8 illustrates the daily temperature and rainfall data from the MSN Weather website [15]. A comparison between the MSN-provided temperature data and the measurements recorded by the deployed monitoring system in Figure 5 indicates a high degree of consistency, validating the reliability of both data sources. This alignment enables a more comprehensive analysis of the relationship between ambient temperature and rainfall at the measurement site. The figure demonstrates a weak correlation between daily temperature and rainfall levels. While heavy rain in early December may have induced minor cooling effects through evaporative processes and atmospheric dynamics, the impact appears marginal. To further investigate the influence of temperature fluctuations on battery performance, capacity retention, and degradation mechanisms, additional high-resolution temperature data is required to refine predictive models and enhance the accuracy of battery health assessments.

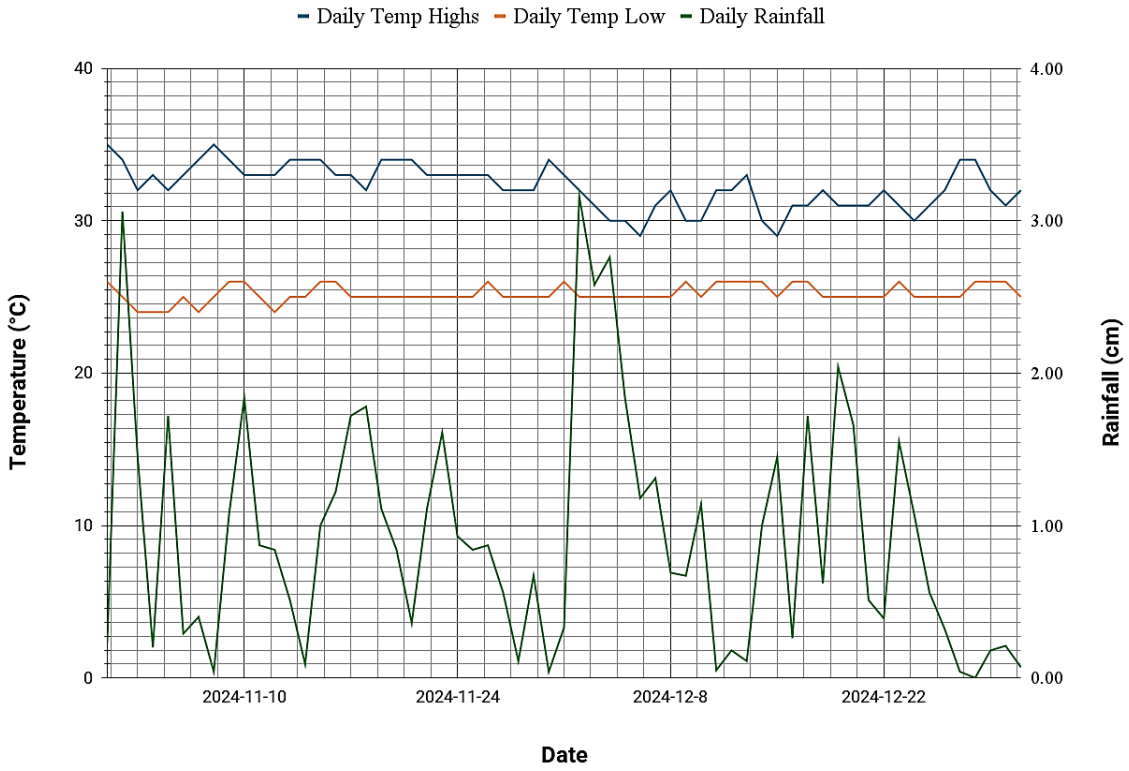


Figure 8. Daily weather record during November - December 2024.

Figure 9 illustrates the relationship between battery voltage and current, with the battery temperature represented through a color map scale. Similar to Figure 7, a notable clustering of data points is observed around 0 amperes within the voltage range of 26 V to 28.5 V, indicating that the battery frequently enters a steady-state condition. This trend suggests prolonged periods of either full charge retention or minimal load conditions, where the battery is neither actively charging nor discharging significantly. Such observations provide critical insights into battery utilization efficiency, depth-of-discharge cycles, and potential long-term degradation trends.

https://doi.org/10.31436/iiumej.v26i3.3612

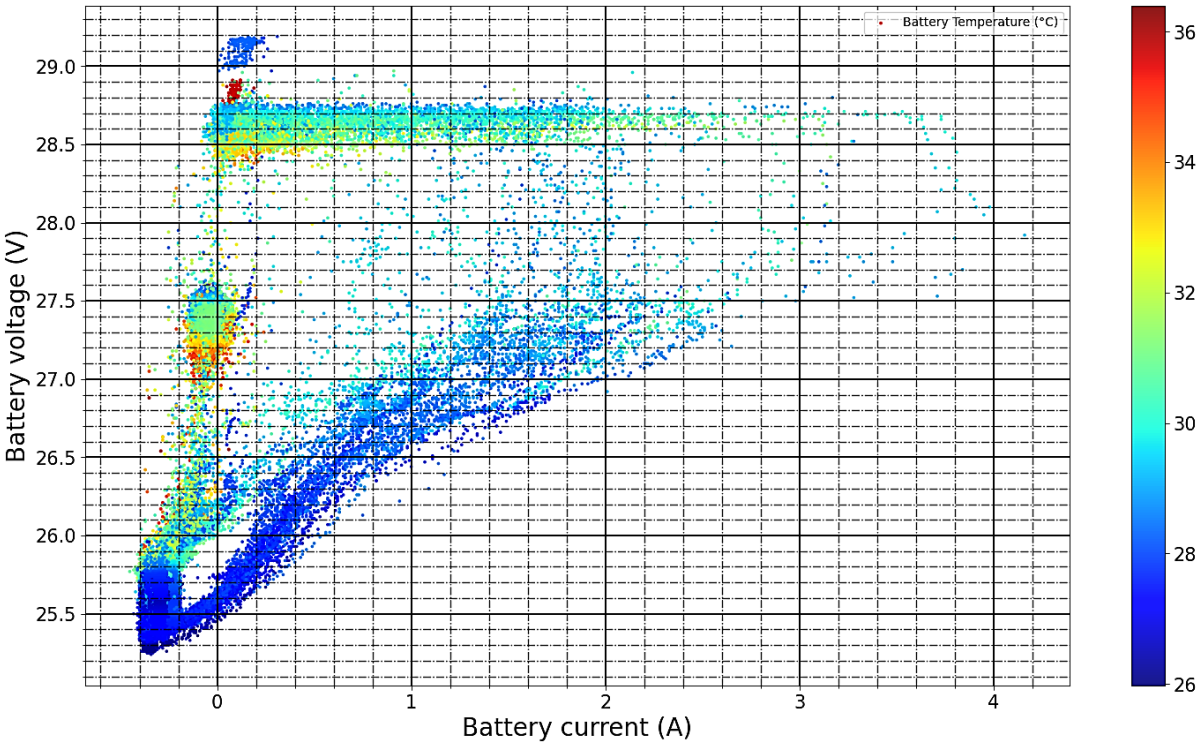


Figure 9. Effect of battery temperature on voltage and current relationship.

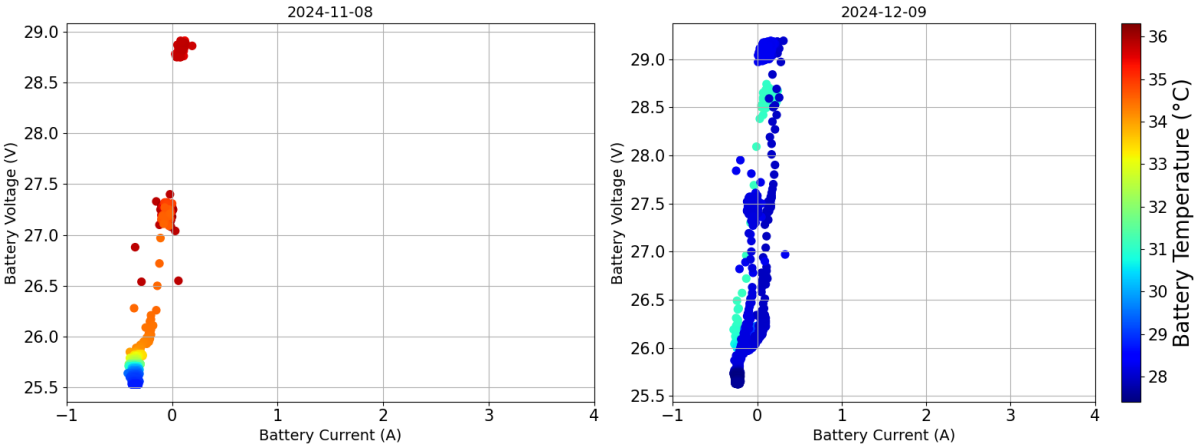


Figure 10. Anomalous data recorded by the monitoring system.

Two significant conditions can be identified and analyzed from the data: the saturation phase of charging and the ohmic relationship at voltages below 28 V.

4.1. Saturation phase of charging

The saturation phase occurs as the battery nears full charge, characterized by a voltage plateau between 28.5 V and 29 V as it approaches maximum capacity. Data plotting reveals a temperature-dependent voltage increase, consistent with theoretical expectations, where higher temperatures correlate with elevated voltage levels due to the impact of thermal variations on electrochemical processes. As shown in Figure 10, anomalous voltage behavior was observed in the data recorded on November 8, where the battery voltage exceeded the expected saturation voltage, likely due to higher operating temperatures. However, a similar anomaly occurred on

https://doi.org/10.31436/iiumej.v26i3.3612

December 9, 2024, where the measured voltage again exceeded the expected saturation voltage despite operating at a lower temperature.

The unexpected behavior observed on December 8 was probably caused by oversaturated charging, which was triggered by intensified electrochemical activity, as indicated by a rapid rise in battery voltage and temperature. This condition activated the controller's protection cutoff mechanism to protect the battery from overheating or further damage, resulting in the load system going offline. During this period, power supply to the load was interrupted, and no data was transmitted. On December 9, internal chemical stabilization may have occurred, allowing the voltage to drop within the operational range. This enabled the system to resume normal operation. However, the anomaly is interesting in that it indicates potential irregularities in internal resistance, warranting further investigation into its root causes and implications for long-term battery performance.

4.2. Ohmic relationship at voltages below 28 V

At voltages below 28 V, an ohmic relationship becomes apparent, predominantly during the initial charging stages and at temperatures below 30 °C. This relationship is defined by a linear increase in voltage relative to current, under Ohmic Law ($V = IR$). The behavior in this region is predominantly influenced by the battery's internal resistance, which causes the voltage to rise proportionally as the current increases. Understanding this region is crucial for analyzing the battery's internal dynamics and efficiency, as excessive resistance can lead to heat generation and a decline in charging efficiency.

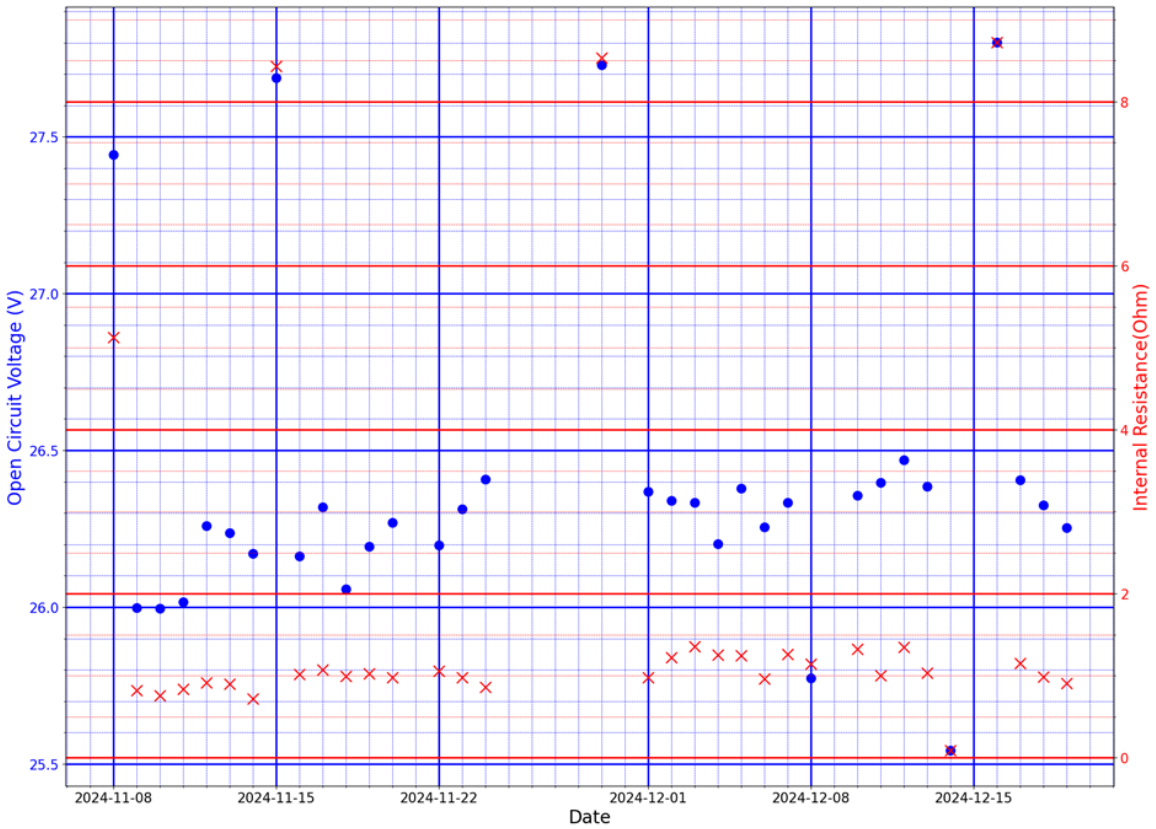


Figure 11. Daily prediction of open circuit voltage and internal impedance of the battery.

Figure 11 illustrates the trends of VOC (represented by blue dots) and IR (represented by red crosses) over time, based on linear regression analysis of current and voltage

https://doi.org/10.31436/iiumej.v26i3.3612

measurements. The VOC estimation was performed using an equivalent circuit model and regression analysis, mapping the voltage-current relationship during battery charging and discharging cycles. The results indicate that VOC remained relatively stable, ranging between 26.0 V and 26.5 V. However, a sudden increase beyond 27.0 V was observed toward the end of the measurement period. This anomaly could be attributed to gaps in data collection on specific dates (e.g., November 8, 15, and 28, 2024) as referenced in Figures 5 and 6, or potential changes in battery operating conditions that were not fully captured.

VOC also indicates the battery's energy-holding capacity, where a decrease in VOC signifies a reduction in available energy storage [16]. Conversely, the ISC represents the battery's ability to deliver electrical current; a higher ISC value indicates better performance in supplying power. ISC displayed a negative mean outflow current from the battery. A more negative ISC value indicates better performance, as the battery degrades, and ISC decreases or is less negative, reflecting a decline in capacity to provide sufficient current [17][18]. This trend is also illustrated in Figure 12, which presents the calculated ISC values, further highlighting the correlation between battery health and current delivery capability.

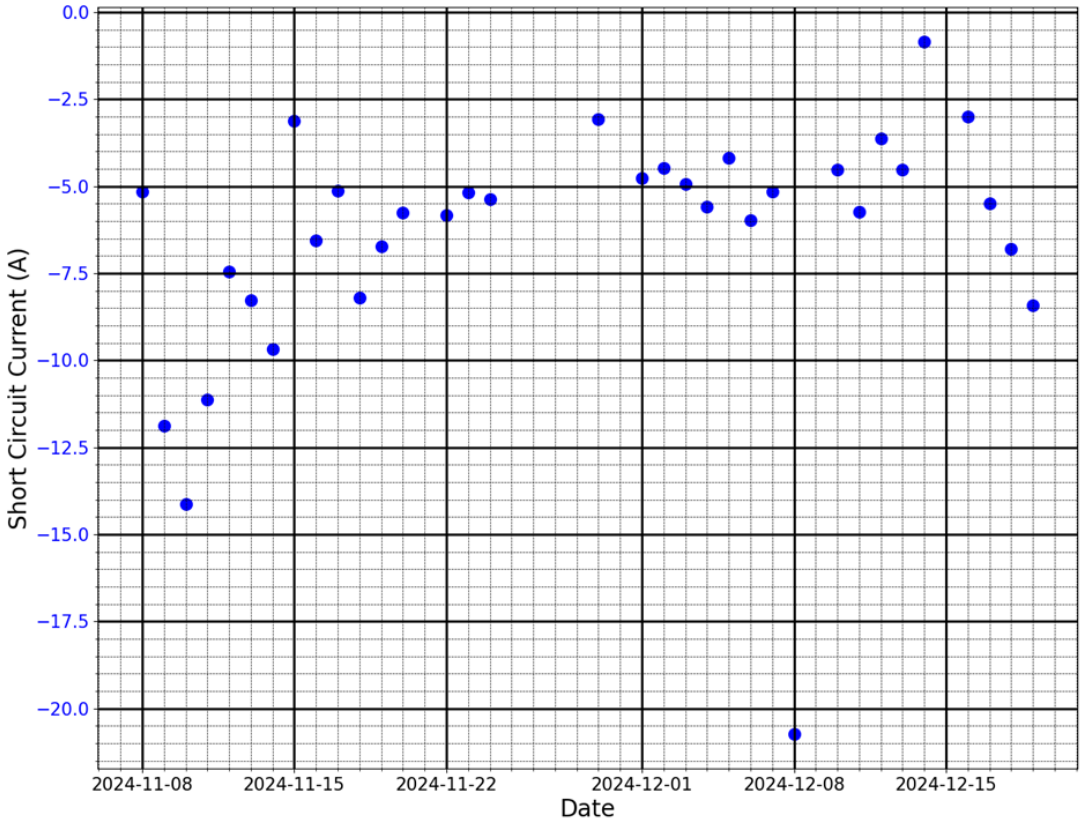


Figure 12. Daily prediction of the short circuit current of the battery.

The internal resistance calculation results align with these findings, showing that IR generally ranges between 0 and 2 Ohms, with a gradual increase over several days, consistent with the observed ISC decline. Interestingly, periodic decreases in internal resistance were also detected, likely due to reversible electrochemical reactions within the electrolyte. In sealed lead-acid (SLA) batteries, oxygen and hydrogen gases generated during charging can recombine to form water, temporarily stabilizing internal resistance [19]. This phenomenon highlights the dynamic nature of battery degradation and the necessity for continuous monitoring. The findings demonstrate that our remote monitoring system effectively tracks and calculates internal impedance, a key parameter for assessing battery health status. Furthermore,

enhancing sensor network efficiency through adaptive data collection and wireless transmission techniques is crucial for overcoming these challenges and ensuring accurate long-term battery health monitoring in solar power stations.

5. CONCLUSION AND FUTURE WORK

This study proved that IoT-based real-time monitoring and predictive maintenance systems effectively maintain lead-acid battery health within off-grid solar power systems, particularly in high-rainfall regions. The developed system provides valuable insights into battery degradation mechanisms by continuously monitoring voltage, current, and temperature. The findings indicate that increasing internal resistance is a key indicator of declining battery performance, affecting the state of health (SOH). Additionally, temperature fluctuations have a significant impact on battery efficiency, underscoring the need for dynamic monitoring and adaptive maintenance strategies. The study showed that integrating predictive analytics in remote monitoring systems can enhance reliability and reduce operational costs. This cost reduction is achieved by lowering labor and transportation expenses for manual inspection and enabling condition-based maintenance, rather than adhering to fixed schedules. Moreover, the research identified challenges in data acquisition due to network disruptions, underscoring the need for more robust infrastructure to ensure consistent, real-time monitoring.

Future research should develop a monitoring system with machine learning integration to improve battery health prediction, while considering computational complexity in resource-constrained environments. It will also enable the examination of the seasonal variations' impact on battery degradation and the exploration of adaptive energy management strategies to optimize battery usage. Additionally, expanding the monitoring framework to support multiple battery chemistries, such as lithium-ion batteries, would provide valuable comparative insights into their performance under similar environmental conditions. This development has substantial implications for real-world deployment, particularly in remote or rural areas where manual maintenance is costly and infrastructure is limited. By enabling autonomous and adaptive monitoring, this system can enhance the reliability of solar-powered installations for off-grid communities, health clinics, and disaster-relief operations. Furthermore, the extension to lithium-ion batteries opens up opportunities for integration into smart grid systems and electric mobility applications, where accurate state-of-health monitoring is critical.

ACKNOWLEDGEMENT

The authors wish to thank the Indonesian Endowment Fund for Education (LPDP) – Ministry of Finance of the Republic of Indonesia for funding this research through the Research and Innovation for Advanced Indonesia (RIIM) program at the National Research and Innovation Agency (BRIN) under grant number B-1468/II.7/FR/10/2022, and PT Jamur Cikuda Nusantara, our research partner, for providing materials and facilities for this research.

REFERENCES

[1] Ahmad Hamdan, Cosmas Dominic Daudu, Adefunke Fabuyide, Emmanuel Augustine Etukudoh, Sedat Sonko. (2024) Next-generation batteries and U.S. energy storage: A comprehensive review: Scrutinizing advancements in battery technology, their role in renewable energy, and grid stability. World J Adv Res Rev, 21(1):1984–1998. https://doi.org/10.30574/wjarr.2024.21.1.0256

[2] Omariba Z, Zhang L, Sun D. (2018) Review on Health Management System for Lithium-Ion Batteries of Electric Vehicles. Electronics, 7(5):72. https://doi.org/10.3390/electronics7050072

[3] Yudhistira R, Khatiwada D, Sanchez F. (2022) A comparative life cycle assessment of lithium-ion and lead-acid batteries for grid energy storage. J Clean Prod, 358:131999. https://doi.org/10.1016/j.jclepro.2022.131999

[4] Singh A, Feltner C, Peck J, Kuhn KI. (2021) Data driven prediction of battery cycle life before capacity degradation. arXiv. https://arxiv.org/abs/2110.09687

[5] Barcellona S, Colnago S, Dotelli G, Latorrata S, Piegari L. (2022) Aging effect on the variation of Li-ion battery resistance as function of temperature and state of charge. J Energy Storage, 50:104658. https://doi.org/10.1016/j.est.2022.104658

[6] Xiong R, Li L, Tian J. (2018) Towards a smarter battery management system: A critical review on battery state of health monitoring methods. J Power Sources, 405:18–29. https://doi.org/10.1016/j.jpowsour.2018.10.019

[7] Selvabharathi D, Muruganantham N. (2023) Real-time monitoring system for lead acid battery health and performance using fuzzy logic and HIL simulator. Int J Eng Trends Technol, 71(7):209–215. https://doi.org/10.14445/22315381/ijett-v71i7p220

[8] Aitio A, Howey DA. (2021) Predicting battery end of life from solar off-grid system field data using machine learning. Joule, 5(12):3204–3220. https://doi.org/10.1016/j.joule.2021.11.006

[9] Li W, Cheng L, Ding W. (2015) Internal resistance of lead-acid battery and application in SOC estimation. In Environmental Science and Engineering. Berlin: Springer; pp 253–260. https://doi.org/10.1007/978-3-662-45969-0_22

[10] Mathew M, Janhunen S, Rashid M, Long F, Fowler M. (2018) Comparative analysis of lithium-ion battery resistance estimation techniques for battery management systems. Energies, 11(6):1490. https://doi.org/10.3390/en11061490

[11] de la Peña Llerandi J, Sancho de Mingo C, Carpio Ibáñez J. (2019) Continuous battery health diagnosis by on-line internal resistance measuring. Energies, 12(14):2836. https://doi.org/10.3390/en12142836

[12] Pham CT, Månsson D. (2018) Optimal energy storage sizing using equivalent circuit modelling for prosumer applications (Part II). J Energy Storage, 18:1–15. https://doi.org/10.1016/j.est.2018.04.015

[13] Łebkowski A. (2017) Temperature, overcharge and short-circuit studies of batteries used in electric vehicles. Przegl Elektrotech, 1(5):69–75. https://doi.org/10.15199/48.2017.05.13

[14] Kim E, Wu B, Shin K, Lee J, He L. (2019) Adaptive battery diagnosis/prognosis for efficient operation. In Proceedings of the Tenth ACM International Conference on Future Energy Systems; 26–29 Jun 2019; New York. New York: ACM; pp 150–159. https://doi.org/10.1145/3307772.3328286

[15] MSN Cuaca. (2025) Prakiraan Cuaca Kota-Tangerang-Selatan, Banten. https://www.msn.com/id-id/cuaca/prakiraan/in-Kota-Tangerang-Selatan,Banten?

[16] Lee S, Kim J, Lee J, Cho BH. (2008) State-of-charge and capacity estimation of lithium-ion battery using a new open-circuit voltage versus state-of-charge. J Power Sources, 185(2):1367–1373. https://doi.org/10.1016/j.jpowsour.2008.08.103

[17] Lai X, Wang S, Liu W, Liu Z, Xu J, Zhang Y, Luo Z, Wang H, Wang Q. (2021) Mechanism, modeling, detection, and prevention of the internal short circuit in lithium-ion batteries: Recent advances and perspectives. Energy Storage Mater, 35:470–499. https://doi.org/10.1016/j.ensm.2020.11.026

[18] Sun S, Xu J. (2024) Safety behaviors and degradation mechanisms of aged batteries: A review. Energy Mater Devices, 2(4). https://doi.org/10.26599/EMD.2024.9370048

[19] Ding Y, Zhang C, Zhang L, Zhou Y, Yu G. (2019) Pathways to widespread applications: Development of redox flow batteries based on new chemistries. Chem, 5(8):1964–1987. https://doi.org/10.1016/j.chempr.2019.05.010

https://doi.org/10.31436/iiumej.v26i3.3615

A DETECTOR FOR TEXTUAL-VISUAL FAKE NEWS USING TEXT SUMMARIZATION AND CONTRASTIVE LANGUAGE-IMAGE PRETRAINING EMBEDDING MODEL

IMAN QAYS ABDULJALEEL[1,2*], ISRAA H. ALI[1]

[1] *Software Department, College of Information Technology, University of Babylon, Hilla, Iraq*
[2] *Department of Computer Science, College of Education for Pure Sciences, University of Basrah, Basrah, Iraq*

**Corresponding author: imanqaysa.sw@student.uobabylon.edu.iq*

(Received: 17 February 2025; Accepted: 29 July 2025; Published online: 9 September 2025)

ABSTRACT: Recently, social media have become an influential means of spreading news in multiple forms, including text, images, audio, and video. With the development taking place today and the heavy reliance of society members on social media to share content online, social media platforms have become a means of spreading unfavorable stories. This study proposes a model to identify false information in various media and determine its authenticity. After pre-processing both textual and visual data independently, the proposed CLIPCrossAttGFN model is used to initialize the fusion feature vector passed to the fully connected layer as part of the classification step. The CLIPCrossAttGFN model starts by passing the image to the Contrastive Language-Image Pretraining (CLIP) visual encoder as a feature extraction method. Also, text, after being summarized to 512 words based on the Bidirectional Encoder Representations from Transformers (BERT), which can identify meaningful words based on the semantic relationship of the text as a whole, is passed to the CLIP encoder before sending the initial extracted features to several layers of CNN, LSTM, and a cross-attention mechanism as a final feature extraction technique. Finally, a multimodal feature is merged using a gate fusion network. The results reveal that the proposed model has better accuracy, ensuring reliable detection compared to other available approaches.

ABSTRAK: Baru-baru ini, media sosial telah menjadi saluran berpengaruh dalam penyebaran berita dalam pelbagai bentuk, termasuk teks, imej, audio dan video. Dengan perkembangan pesat berlaku hari ini, dan kebergantungan tinggi masyarakat pada media sosial bagi berkongsi konten secara atas talian, platform media sosial telah menjadi medium penyebaran maklumat yang tidak sahih. Kajian ini mencadangkan model bagi mengenal pasti maklumat palsu dalam beberapa media dan menentukan sama ada ianya benar atau palsu. Selepas kedua-dua data teks dan visual melalui proses pra-memprosesan secara berasingan, model CLIPCrossAttGFN yang dicadangkan diguna pakai bagi memulakan vektor ciri gabungan yang dihantar ke lapisan bersambung sepenuhnya sebagai langkah pengelasan. Model CLIPCrossAttGFN bermula dengan menghantar imej kepada pengekod visual Contrastive Language-Image Pretraining (CLIP) sebagai kaedah pengekstrakan ciri. Teks pula diringkaskan kepada 512 perkataan berdasarkan Perwakilan Pengekod Dwi Arah daripada Transformers (BERT), yang dapat mengenal pasti perkataan bermakna berdasarkan hubungan semantik teks secara keseluruhan, dihantar kepada pengekod CLIP sebelum ciri awal diekstrak ke beberapa lapisan CNN, LSTM dan mekanisme silang perhatian sebagai teknik pengekstrakan ciri akhir. Akhirnya, ciri multimodal digabungkan menggunakan rangkaian gabungan berpintu. Dapatan kajian mendapati bahawa model yang dicadangkan mempunyai ketepatan tertinggi, menjamin pengesanan yang boleh dipercayai berbanding pendekatan sedia ada.

https://doi.org/10.31436/iiumej.v26i3.3615

KEYWORDS: *False news, Deep learning, Gate fusion, Multimodal data, BERT.*

1. INTRODUCTION

To simplify formatting, you can use the style menu located just below the standard menu. Every button has a name similar to the style name in brackets after each paragraph. (Normal). In the modern era, social media has spread terrifyingly, leading to the almost complete abandonment of other means of transmitting information, such as television, newspapers, and magazines. Therefore, any individual can be a source for spreading news, which may often be fake, to achieve a specific societal, political, or religious goal.

Fake news can be defined as misinformation disseminated through the deliberate imitation of prevailing news, distorting and fabricating it to make it appear real. Fake news often relies on real news after injecting it with false information, according to the intended purpose. This modification affects not only texts but also images and videos. Therefore, spreading fake news through social media has become a frightening threat to our real society. Fake news recorded its highest levels from 2020 to 2021, which coincided with the COVID-19 pandemic, immediately linked to quarantine and total reliance on social media as a source of information. Fake news began to be used after two political events: the 2016 US elections and Brexit [1].

The most famous social media on the Internet are Facebook, Twitter, Instagram, and TikTok, which have become indispensable sites for many individuals and companies to publish their news on the one hand and promote their businesses on the other. However, this freedom to publish content may lead to the publication of misleading content that relies on different multimedia, such as texts and images, to attract users and promote the publication of fake news. Because of this, detecting fake news becomes a big challenge for everyone to confront. For this reason, many algorithms built to detect fake news based on data sets from social media platforms: the first of which relies on the content of the published news to determine its characteristics according to the textual, image, or video content, while the other based on the social context, which based on the user's profiles on social media and his activity in responding, the activity of the post, how it is published and expanded, and the structures of dissemination to conclude the validity of the published news [2].

One of the primary reasons for analyzing news and identifying fake news is that some individuals or organizations, more specifically, seek to exploit the spread of fake news to achieve personal or political gains. Others rely on spreading fake news to mislead people and make them adopt false beliefs to achieve goals that destabilize and dissolve society at all levels. Others seek to erode people's confidence in the transmitted news, often reducing confidence in the country's governments and political leaders, thereby undermining security [3].

Deep learning (DL), a subset of artificial intelligence that deals with complicated problems and big datasets by automatically collecting the characteristics of multiple layers of neural networks in a short time, has grown in favor of various single- and multimodal false news detection models [4]. The most often used neural network models are recurrent neural networks (RNNs), convolutional neural networks (CNNs), and attention networks. Using these models to identify the distinctive features of fake news, whether its content is textual (Natural Language Processing (NLP)) or visual (Computer Vision), has yielded impressive detection results [9] by identifying special and different features according to the content. However, as most news is primarily text-based, text remains the primary source for identifying fake news features despite the limitations of text analysis models and their problems in capturing long-term dependencies between words in a single text to create improved textual representations of

sentences [5]. However, sources that rely on multiple content, whether text and visual or text and video, are still few compared to research that relies on a unimodal. The main reason for the lack of work in the field of multimodal fake news detection is the difficulty of choosing methods for integrating multimodal content and the effectiveness of their implementation because most models fail to select accurate individual features specific to each pattern and contribute valuable information to detect fake news and exclude noise resulting from the fusion of multimodal features., which makes it one of the most significant challenges in this field [1]. The increasing diversity of fake news on social media, along with its coverage of various topics, poses a substantial challenge to those working in this field, which is often characterized by media diversity. Therefore, classifiers must choose a method to extract essential features from this multimodal data, whether visual or linguistic. Through numerous studies, relatively simple models have emerged that rely on encoding images and texts based on various models, each tailored to the medium's environment, followed by different fusion methods, such as cross-modal attention and concatenation. Recently, several models have emerged that analyze images and text as a linked pair to reveal the relationship between them, including the Contrastive Language-Image Pretraining (CLIP) and Bootstrapping Language-Image Pretraining (BLIP) models [6].

While various models focusing on language and vision harmony, such as the Vision-and-Language Bidirectional Encoder Representations from Transformers (ViLBERT) model, which uses Faster Region-based Convolutional Neural Network (R-CNN) to determine regions of interest in the image and BERT for text, it works by dividing both image and text handling into two separate channels. Then, Co-attention is used to merge the output of both channels. This structure is frequently employed in visual question answering because it works effectively in problems requiring a deep comprehension of visual content. The Universal Image-Text Representation (Uniter) model pre-combines the inputs from both channels, text, and areas of interest to generate a data vector. In contrast, ViLBERT preserves the independence of both text and image channels before building the correlation process. The BERT model receives this data vector and uses it to identify deep correlations between the inputs. So, it is perfect for text-to-image matching [7]. CLIP is one of the most important models for understanding multimodal data features. It is a learning model developed by OpenAI to understand the relationship between the contents of a digital image and its corresponding text words using contrastive learning on multiple pairs (text, image). Its work is defined by encoding the text and visual data in interconnected vector representations, thus enabling it to find the semantic relationship of its multimodal inputs with a high level of understanding [8]. The benefit of the CLIP model is that it offers a thorough understanding of the level of text-image consistency overall, sometimes eliminating the need to look for forgeries in individual visual objects or textual details as in ViLBERT or Uniter models. Thus, its use frees it from the difficulties of internal consistency of each text and image component, allowing for speed of development and simplicity of structured representation [8].

There are two main types of fake news detection methods: unimodal and multimodal. The first method detects bogus news using a single information, such as text, image, audio, or video. As a result, these methods rely on only a limited portion of the available information, restricting their general understanding of the news story and determining their capacity to distinguish between truth and falsehood. The second type overcomes the shortcomings of the first by detecting fake news using accessible data from various linguistic and visual domains. Despite the superiority of multimodal detection methods in detecting fake news, they are still confronting several challenges, the most important of which is the mechanism for integrating information from various media to benefit from the most significant number of features that distinguish the truth of the news from its falsehood. The third issue is a shortage of multimodal

https://doi.org/10.31436/iiumej.v26i3.3615

data sets that include various information types like text, image, video, and audio. This section will review the most relevant connected papers dealing with these two methodologies. In the paper by Kaliyar [9], the textual fake news framework structure is built on the BERT coding paradigm and utilizes convolutional layers to extract key text properties. Based on the Fake News dataset, the model accuracy is 98.9%. Rustam applied Artificial Intelligence (AI) mechanisms, including machine learning (ML) algorithms and deep learning, to a system for identifying false news. This approach converted textual news input into a visual representation, collecting accurate and crucial features that enhance fake news detection accuracy. The accuracy rate of 99.7% using the ISOT dataset, but the LIAR dataset is below 92% [10]. Work by Daik in [11] constructed a fake news detection system using a convolutional neural network. Still, this system struggled to collect enough image samples to train the model and maintain high accuracy.

In [12], complementary attention fusion is combined with a deep neural network to detect the associated features between text and image in the input dataset. The deep neural network consists of three fully connected layers (FCL) to extract high-level features that significantly detect fake news based on the annotations and texts extracted using CAF. The datasets used are GOSSIPCO, POLITIFACT, FAKEDDITE, and PHEME, and the model obtained an accuracy of 85%, 90%, 95%, and 90% respectively. In [13], the authors proposed a model that extracts text features using a multi-level encoding network, while image features are extracted using the VGG-19 network. Then, the resulting text and image features are passed into a relationship-aware attention network that can detect the similarity between information pieces in different media. Finally, the resulting fusion features are fed into a fake news detector. This approach is applied to three reference datasets: Weibo, Twitter, and PHEME, and the accuracy rates were 90%, 85.5%, and 87%, respectively. A multimodal detecting fake news model is presented in [8] based on contrastive learning. The model begins by extracting essential features from the input text and image at various levels of granularity through contrastive learning. Then, a multi-granular representation of multimedia news is applied, utilizing both image and text features in parallel. The model determines whether the news is real based on multi-head self-attention techniques. The RECOVERY, GOSSIPCOP, and MR2 datasets are used. The accuracy of each of the above datasets was 85.48%, 92.3%, and 89.45%, respectively. The research in [14] presents a method based on contrastive learning using CLIP and semantic alignment, applied to textual and image information, utilizing three datasets: PHEME, GOSSIPCOP, and MR2. The accuracy on each dataset was 91.29%, 91.98%, and 89.86% respectively.

The research [15] relies on quantum fusion in detecting multimodal fake news. Text features are extracted using Transformer-XL Networks (XLNet), and image features are extracted using VGG-19. Then, text and image features are combined and passed to a proposed quantum convolutional neural network to detect the truth of news from its falsity. Using two datasets, Gossip and POLITIFACT, with accuracies of 87.9% and 84.6%, respectively, we can observe the model's ability to detect and distinguish fake news. In [16], a multimodal news model relied on the representation of BERT to extract the text characteristics and used VGG-19 to extract the image characteristics. In addition, a multimodal compact bilinear pooling technique is used to merge textual and image attributes with an attention mechanism, resulting in one time and integrating textual and image attributes independently at another time. This proposed model was applied to two data groups, Twitter and Weibo, with accuracies of 68% and 81%, respectively. In the paper by Rashid in [17], a model based on a robustly optimized BERT approach (RoBERTa) is used to analyze the entrance text, and ResNet-50 is used to analyze the images associated with it as a first stage in the false news detection system which received an accuracy of 91 % using a Twitter dataset and 93 % using the Weibo dataset.

The main contributions of this research have been outlined as follows:

It offers a novel strategy for detecting multimodal false news based on the CLIP model, primarily for text and image pairs. This is followed by numerous parallel or sequential layers of Convolutional Neural Network (CNN), Long-Short-Term Memory (LSTM), and attention mechanisms.

1. Uses several techniques to clean the text by removing stop words and roots and adopting the spelling correction of words by 1-2 letters to retain the possible words distorted inadvertently.
2. Use the word summarization technique to extract the most important words in the text that are longer than 512 words. Adopt the BERT approach to summarize texts and detect the associations between words to determine their importance within the text.
3. The suggested architecture's efficacy is evaluated by accuracy, precision, recall, and the F1 score on two datasets: GOSSIPCO, RECOVERY, and Ti-CNN.

This paper has the following structure: In the second section, we give the background material used in this paper. Section three covers the proposal model by detailing the components of the suggested model. The fourth component covers the dataset employed, discusses the results of the proposal model, and compares it to other models. The fifth part described the work's results and the most crucial future ideas for overcoming the problems we encountered while working.

2. MATERIAL AND METHODS

2.1. Contrastive Language-Image Pretraining Embeddings Model

The proposed model begins by utilizing a pre-trained characteristics unit for managing multimodal data, termed CLIP, which combines text and image information to create a feature space. The CLIP relies on the transformer structure and operates on the principle of "Comparative Learning" based on both contrastive loss functions, to achieve a thorough grasp of the image and text through large-scale pre-training [11]. Thus, CLIP seeks semantic similarity between its two interrelated textual and visual inputs rather than precise information, distinguishing it from other models, particularly those dealing with multimodal inputs [9].

The CLIP model structure is divided into the image encoder and feature extraction, which are generated with pre-trained convolutional neural network models like ResNet-50 or Vision Transformer (ViT). The other part is for the text encoder, which is based on transformer principles such as BERT and Generative Pre-trained Transformer (GPT). In the training stage, input pairs (text-image) are matched to increase the cosine similarity of the matched pairings while reducing the cosine similarity of all other non-matched pairs. To minimize retraining on the subsequent dataset, the CLIP team collected 400 million and 1 billion matched pairs of images and texts for training. Since CLIP is pre-trained on some datasets like ImageNet, the text encoder can recognize rich natural language semantics, making it suitable for zero-shot prediction [11].

Zero-Shot Learning (ZSL) is a challenging machine learning paradigm [6]. Its goal is to recognize previously unseen objects or concepts. Traditional models require extensive datasets for each expected category. In contrast, ZSL uses supplementary data, such as semantic attributes, to infer the properties of unseen categories. The CLIP model’s ability to embed images and text into a shared space naturally extends this concept. It enables effective

https://doi.org/10.31436/iiumej.v26i3.3615

knowledge transfer from seen to unseen categories [6]. The structure of CLIP comparative learning is visible in Figure 1.

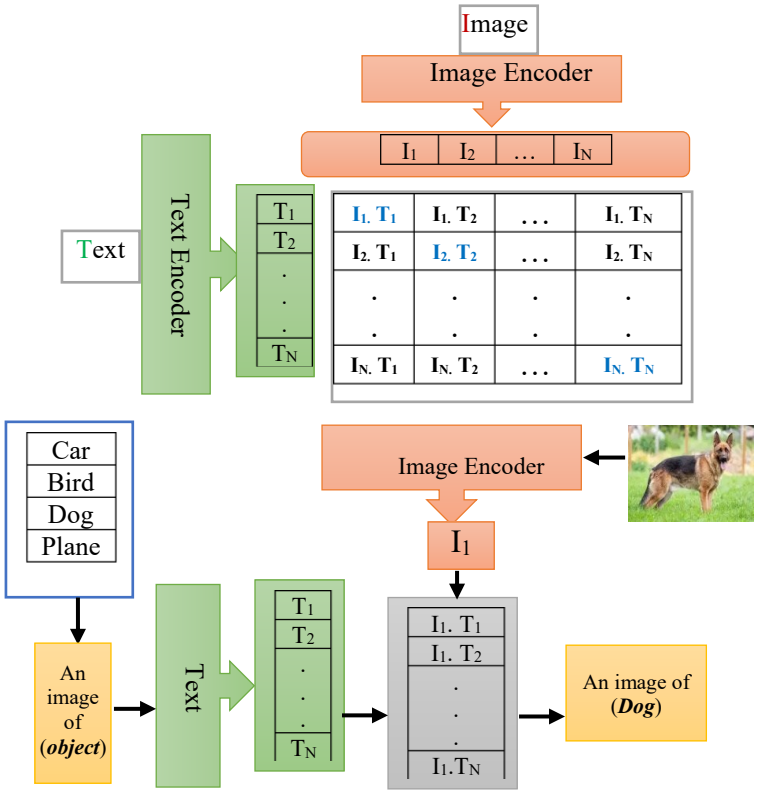


Figure 1. Structure of CLIP Contrastive Learning [6]: (a) Contrastive pre-training, (b) Dataset generation based on "text" and zero-shot prediction.

2.2. Cross-attention Mechanism

In recent years, attention mechanisms have gained popularity in various fields, although they originated in natural language processing, specifically in machine translation. The model by Bahdanau et al. [18] presented at that time could find all the information related to the specific words in the source sentence. Two popular types appeared in the multimodal datasets (image-text) models: co-attention and cross-attention. These two models could capture the unimodal and multimodal relationship with each other, and thus the ability to discover important features and detect fake news [13]. The cross-attention module is based on the self-attention module in its construction, which uses two inputs with different sequences (such as text and image), so the attention score is calculated based on both formats. Figure 2 describes the Architecture of the cross-attention mechanism.

The image-text features fusion process starts by initializing two vision feature vectors (VF) and one text feature vector (TF) as inputs to the cross-attention network to obtain text attention features (TFa). Also, the vision attention features (VFa) are obtained by passing two text feature vectors (TF) and one vision feature vector (VF) to the cross-attention network. They can be expressed mathematically by the following two equations, as in Eq. (1) and Eq. (2) [19] where TL is the transform layer, (Q, K, V) are the attention input, and $\sqrt{p}$ is the length of the aligned vectors.

$$TF_a = SoftMax\left(\frac{TL_Q(VF).TL_k(TF)^T}{\sqrt{P}}\right)TL_v(TF) \quad (1)$$

https://doi.org/10.31436/iiumej.v26i3.3615

$$VF_a = SoftMax\left(\frac{TL_Q(TF).TL_k(VF)^T}{\sqrt{P}}\right)TL_v(VF) \quad (2)$$

The outputs of the two previous equations are then passed to a fully connected layer to process and improve the resulting representation. After that, layer normalization is applied to enhance the network's stability during training.

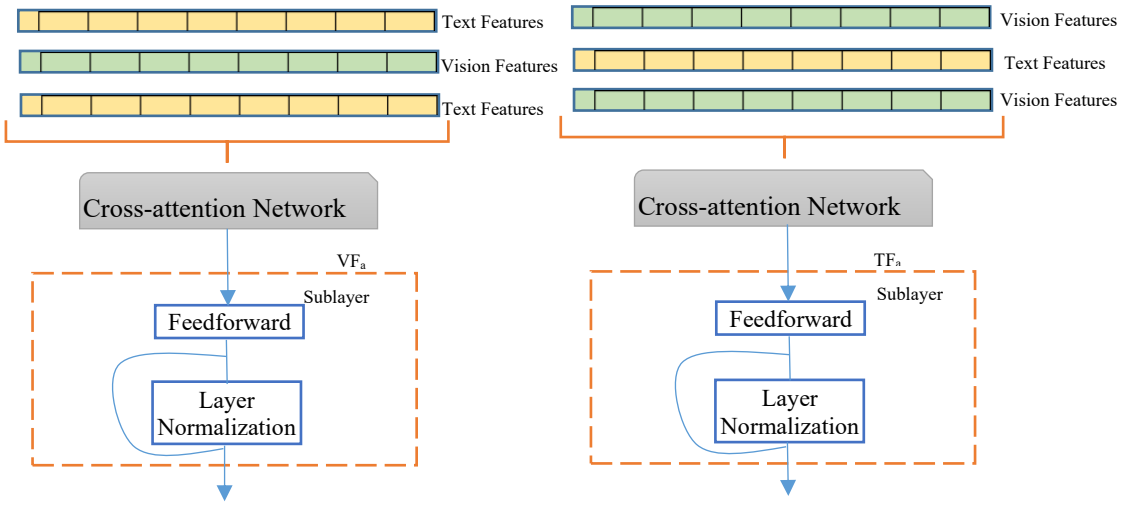


Figure 2. Architecture of the Cross-attention mechanism [19].

2.3. Gate Fusion Network (GFN)

A gate fusion module is used to understand the extracted features and how to refine them by removing noisy modality inputs that negatively affect fake news detection in a multimodal environment [6]. Its inputs are text and image extracts from previous stages, which control the selection of specific features by this gate fusion. There are several ways to combine gates described as follows [20]:

1. Independent Gates (IG): Two sets of gates are defined, one for extracting text features and the other for extracting image features. Each gate loads information specific to only one typical feature. In this paper, we used this type.
2. Cross-Gate (CRG): Here, two sets of gates are also defined, but here the two gates are multiplied by other feature maps, which increases the learning opportunity between multimodal patterns.
3. Complementary Gates (CG): Here, the connection between channel weights before the text and image feature fusion stage, and the complementary features are extracted based on the initialization of two separate gates. Here, it is evident that the two gates learn complementary information from the same concatenated weights.

3. METHODOLOGY

The work begins by preparing dataset totals in a way that the model can deal with in its bilateral form (text, image). Initially, text and image are processed separately by cleaning the text data and removing stop words before all words are returned to their root form using stemming and lemmatization. The long sentence text input is summarized at an optimal length before text embedding using CLIP. The image is resized to 224×224 pixels before being used with CLIP for feature extraction. After that, maintain the text linked with each image in the dataset to assist in dealing with it later using the proposed "CLIPCrossAttGFN" model to

https://doi.org/10.31436/iiumej.v26i3.3615

conclude the work with a result that distinguishes between true and disinformation. Figure 3 explains the general structure of the proposed model. Figure 4 describes a general diagram of the proposed multimodal fake news detection framework.

3.1. Image and Text Pre-processing

The dataset is pre-processed based on its type, where image data is analyzed by verifying the validity of image download links (URLs), then downloading images and resizing them to a uniform size of 224×224 pixels. As for text data, pre-processing involves removing all symbols, stop words, punctuation marks, links, and emojis to keep only words as a basis for assessing the importance of the existing text. After that, make all words lowercase, then root words by removing prefixes and suffixes to return them to their original form. Additionally, an attempt is made to correct the spelling of words to the minimum number of letters (two letters maximum) to retain the highest possible number of linguistically understandable words, which appears to be due to human error.

3.2. Text Summarization

To understand natural languages, models have recently emerged that can understand human languages, specifically using deep learning models. One is the BERT (Bidirectional Encoder Representations from Transformers) model. This model segments the input text into words and further segments the words into sub-words to understand the exact meaning of these smaller parts, which often include some that are not commonly used. BERT adds two tokens at the beginning and end of the input sequence to help better understand the language: CLS (at the start of the text input sequence) and SEP (at the end of the text input sequence). It then uses segment embeddings to identify the most distinctive words based on contextual information throughout the text, thus providing a comprehensive understanding before selecting the most important words [9]. After cleaning the pre-processing text, work on summarizing it to get the most influential words in the entered text with lengths compatible with what can be embedded using the CLIP model. For this, it is working on passing the words in the text to the BERT embedding technique, which calculates the weights of each existing word and the number of times it is duplicate, then choosing the most important ones based on the highest weight while allowing the word repeating in a proportion that is compatible with the number of times it is duplicate without neglecting other import words. Another problem is the appearance of unknown words in BERT. To address this, they are divided into sub-words to find similar words in BERT. However, some sub-words are converted into special symbols (like ##) because they are not found in BERT. Therefore, in the event of a similar situation, the parts of the word are re-merged and returned to their original state and dealt with as a new word with a specific weight.

3.3. The Proposed Model (CLIP-Cross-attention-GFN Model)

To deal with textual and visual news inputs, we aim to utilize transformers to capture the interactions within the patterns, whether they are textual or visual information. For textual data, it uses the pre-trained CLIP embedding (encoder) for texts to search for semantic information in the input news text words. As for image inputs, we also use the CLIP encoder, but we extract the image features of the news in the part related to dealing with visual information. For the text embeddings generated by a pre-trained CLIP model, we get 768-dimensional word embeddings, while for image embeddings, we also get 768-dimensional embeddings using the CLIP model.

https://doi.org/10.31436/iiumej.v26i3.3615

The text features embedded using CLIP are passed into two models in parallel to obtain the spatial features of specific words, which indicate homogeneous meanings, such as a negative sensation. This can be obtained by concurrently sending the text features utilized via CLIP into multiple models. In addition, the second parallel model consists of two sequentially ordered layers of LSTM, which are applied to gather the text's temporal properties and contextual coherence from start to finish. Then the outputs of the two previous models (CNN and LSTM) and the text features embedded using CLIP are passed to a dense layer (a set of regular cells in a neural network to prevent overfitting). While the image features are embedded using CLIP, which are also passed to another dense layer, before the outputs of the text and image dense layers are passed to a cross-attention mechanism.

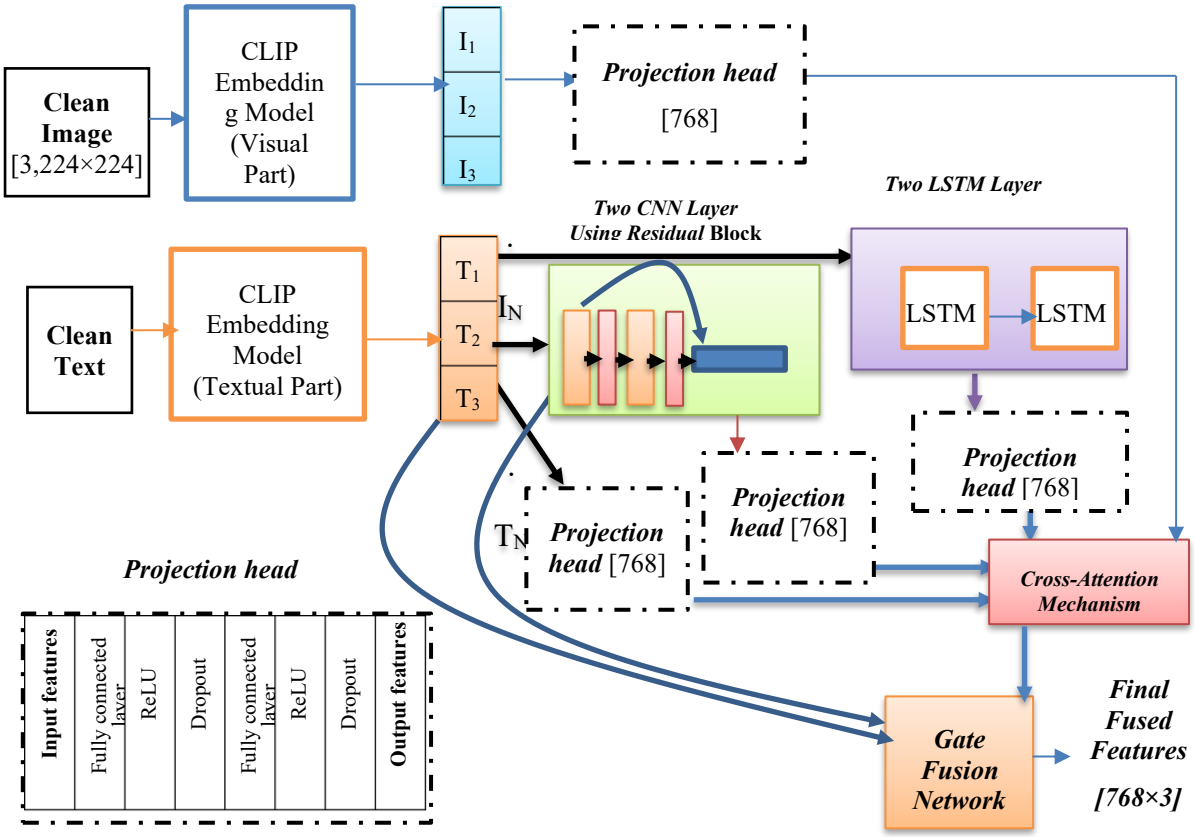


Figure 3. The proposal of the CLIPCrossAttGFN model.

Up to this stage, the most important features will be extracted. To obtain other features that we may not have been able to extract using the previous steps, we are now working on passing the outputs of the cross-attention mechanism in addition to the text features extracted using CNN and the text features embedded using CLIP as inputs to the GFN network to select the features based on the most influential ones, whether textual or visual. The final vector of features for this stage represents the vector of the extracted, combined textual and visual features. Leveraging DL methods at the appropriate locations to extract the most essential features during the next step gives the CLIPCrossAttGFN model its efficacy. First, it uses the pre-trained CLIP model, which begins by thoroughly depicting the semantic link between two input modalities (textual and visual). Second, it uses CNN and LSTM models to improve comprehension of information according to the news's structure and context. Finally, it uses a GFN network that retrieves the feature vector without distortion using cross-modal attention.

https://doi.org/10.31436/iiumej.v26i3.3615

3.4. Classification Model

After merging the textual and visual features based on the CLIPCrossAttGFN model, we now rely on a fully connected layer, followed by a special function to predict the validity of the news with its two probabilities (True or False), which is the Sigmoid function. We split the fake news datasets in a ratio of 8:1:1 for training, validation, and testing. The number of LSTM neurons is 768, while CNN filters and kernel windows are 768 and 3, respectively. The following hyperparameters are organized to improve model training and testing: AdamW, initial learning rate, weight decay, number of epochs, batch size, and dropout. Their values are entered manually according to Table 1.

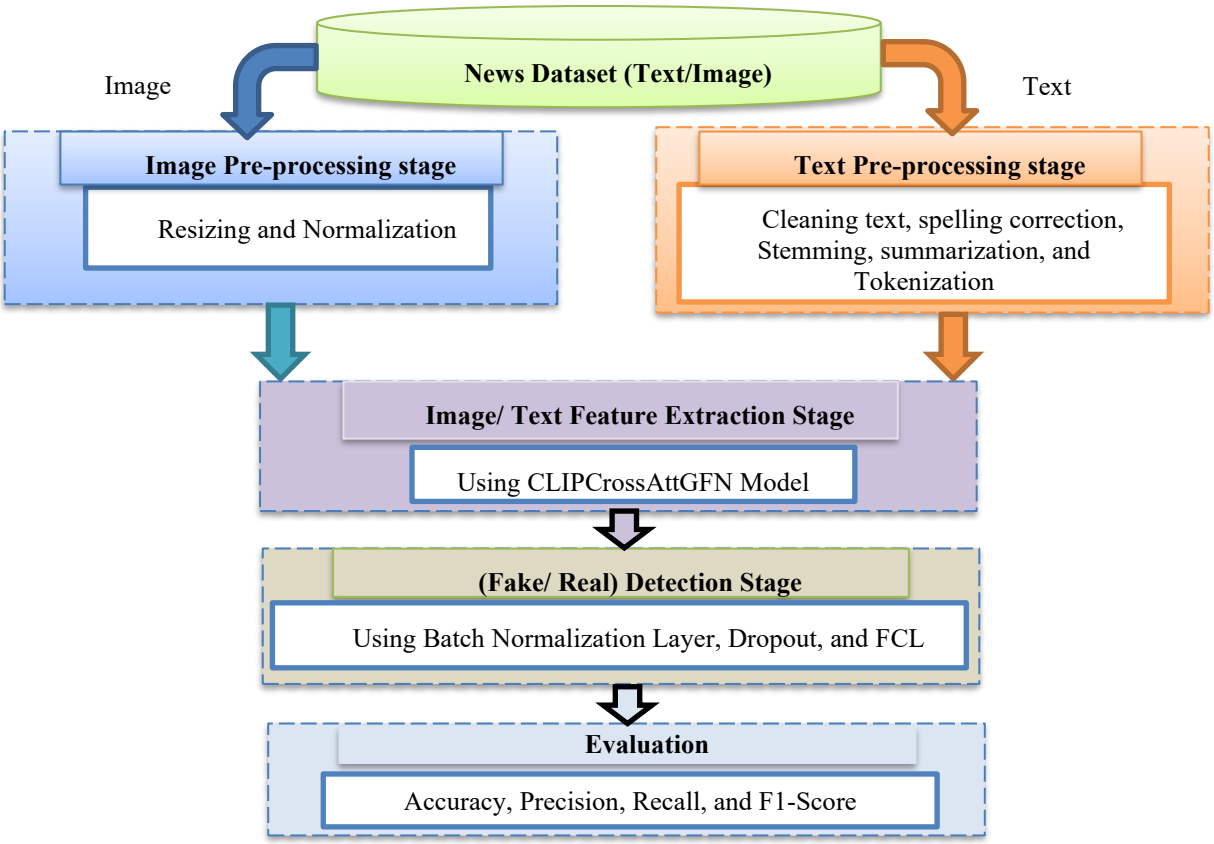


Figure 3. The general diagram of the proposed multimodal fake news detection Framework

Table 1. Test configuration of parameters following improvement

Hyperparameters	Value	Hyperparameters	Value
Optimizer	AdamW	Number of epochs	20
Initial learning rate	1e-4	Batch size	64
Weight decay	0.01	Dropout	0.5

3.5. Evaluation Metrics

Several evaluation metrics are used to demonstrate the efficiency of the proposed fake news model, the most important of which are the accuracy metric, followed by the precision, recall, and the F1 score. Accuracy alone may provide unrealistic results if the input data is unbalanced. The four metrics can be represented as follows in Eq. (3) to Eq. (6) [2].

https://doi.org/10.31436/iiumej.v26i3.3615

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \times 100 \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \times 100 \quad (5)$$

$$F1 - Score = 2 \times \frac{Recall \times Precision}{Recall+Precision} \quad (6)$$

4. EXPERIMENTAL RESULTS

4.1. Fake News-related Datasets

The GOSSIPCOP dataset from FakeNewsNet is the first dataset we use, which is based on English-language text websites and classified into two categories: real (label = 1) and fake (label = 0) [12, 8]. We also used Megan Risdal's Fake News dataset in Kaggle (Ti-CNN). It contains news articles of varying lengths, ranging from 5 to 3271 words, and URLs to images; unfortunately, some are corrupted due to the closure of these sites and their subsequent inoperability. Table 2 shows the number of fake and real news [8]. The third dataset we used is RECOVERY, which is considered one of the most essential datasets containing healthy multimodal information. Its articles were collected during the COVID-19 pandemic (i.e., during the first five months of 2022) based on several reliable and unreliable websites, a total of which reached 60 websites [12].

Table 2. Test configuration of parameters following improvement

Dataset	GOSSIPCOP dataset			RECOVERY			Ti-CNN		
	Train	Validate	Test	Train	Validate	Test	Train	Validate	Test
Separated size	10104	1263	1264	1142	143	143	2825	353	354
Total size	12631			1428			3532		
Maximum length of text	9137			9983			3271		
Class	2			2			2		

4.2. Weight Adjustment

Many multimodal fake news datasets suffer from an imbalance in the main category (like real and fake labels) due to a big difference in the number of samples for the categories that make up each dataset which done sometimes due to the damage to one item in the input pair (text, image), in addition to relying on a specific type of news (such as sport news, medical news, and art news) that is likely to contain a particular category less than the other category. During model training, these reasons led to the model being biased towards the most numerous categories. There are many ways to address the imbalance in the datasets, the most important of which are: oversampling (repeating samples for the least multiple categories in the dataset) and undersampling (reducing samples for the most abundant categories in the dataset). In this research paper, we used another type called "class weight adjustment", which depends on assigning different weights to samples from various categories, which in this way allows the model during training to focus on the categories that have the least number of samples and thus gives good results in its ability to learn from them [21].

https://doi.org/10.31436/iiumej.v26i3.3615

4.3. Performance and Comparison Results

In this section, we evaluate the model's performance and compare its results with state-of-the-art models. Table 3 compares the performance of various networks for multimodal fake news identification. Existing approaches such as BERT-VGG16, BERT-VGG19, and BERT-ResNet50 are evaluated and compared to the suggested design. The CLIP for text-image data produces more accurate findings. The achieved outcomes based on the performance metrics: accuracy, precision, recall, and F1-score, with corresponding values of around 86%, 85.37%, 86%, and 85.57% in the GOSSIPCOP dataset. The results for the performance measures for the RECOVERY dataset are 90.21%, 89.58%, 90.21%, and 89.74%, whereas in the Ti-CNN dataset, they are 95.47%, 95.52%, 95.47%, and 95.4%, respectively. A comparative analysis was carried out using the dataset and considering the results obtained from the various performance indicators. Table 4 shows an analysis of the proposed technique and earlier models.

The proposal detection results based on the multimodal fake news model are presented for comparison with BERT-VGG16, BERT-VGG19, BERT-ResNet50, and BERT-EfficientNetB0 models, as well as with other research papers [1, 2, 12, 22, 23, 34], as explained in Tables 3 and 4. Datasets containing texts exceeding three thousand words were relied upon to test the ability of the proposed summarization technique based on BERT embedding to reduce text length before merging them with the associated images and sending them to the proposed fake news detection model. We also note that the data for the two patterns, real and fake, is unbalanced, which requires "Class Weight Adjustment" to help deal with this dataset. Another limitation is that this proposed model works only on text and image datasets, and other data types, such as video and audio, cannot be used because the CLIP embedding model currently available only deals with these two media (text and image). In addition, the currently available datasets are relatively small because they are derived from datasets built at previous times, which resulted in the loss of some of them due to the interruption of their sites. Although the proposed framework performs effectively, Table 5 shows that it occasionally fails to obtain accurate categorization. Furthermore, the nature of some previously pre-trained emotional phrases and images that the CLIP model perceives as real is very similar to that of some false news generated with the same emotional formula and visuals, making it more likely to be regarded as real at the news input level. Another factor is the imbalance between the quantity of false and true news articles, as seen in Table 2, which leads to poor generalization.

Table 3. Performance analysis of proposed models with various multimodal models

Models		**BERT-VGG16**	**BERT-VGG19**	**BERT-ResNet50**	**BERT-MobileNetV2**	**Ours Model**
RECOVERY dataset	**Accuracy**	74.13	73.43	79.02	81.12	90.21
	Precision	80.02	81.50	83.85	84.53	89.58
	Recall	74.13	73.43	79.02	81.12	90.21
	F1-score	76.58	76.50	80.90	82.49	89.74
GOSSOPCO dataset	**Accuracy**	61.23	62.10	68.32	75.14	86
	Precision	66.44	67.57	71.70	76.93	85.37
	Recall	61.23	62.10	68.32	75.14	86
	F1-score	63.48	64.18	70.36	76.88	85.57
Ti-CNN dataset	**Accuracy**	87.25	87.91	88.12	92.07	95.47
	Precision	87.55	87.73	89.86	92.64	95.52
	Recall	87.25	87.91	88.12	92.07	95.47
	F1-score	87.36	87.98	89.15	92.20	95.40

https://doi.org/10.31436/iiumej.v26i3.3615

Table 4, Comparison evaluation of proposed and current methods for multimodal fake news

References	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score
[22]	RECOVERY	79.03	72.51	45.76	56.11
	GOSSIPCOP	85.69	84.62	75.12	79.59
[12]	GOSSIPCOP	86.3	87	95	90.5
[23]	GOSSIPCOP	84.8	82.2	79.7	80.8
[24]	RECOVERY	81.72	72.68	66.41	69.40
	GOSSIPCOP	88.04	86.77	80.01	83.25
[2]	GOSSIPCOP	74.8	74.87	75.31	74.71
[1]	Ti-CNN	95	97	96	96
Ours Model	RECOVERY	90.21	89.58	90.21	89.74
	GOSSIPCOP	86	85.37	86	85.57
	Ti-CNN	95.47	95.52	95.47	95.40

Table 5. Samples of real and outcome labels

News URL	Downloaded Image	Title	Real Label	Outcome Label
www.dailymail.co.uk/tvshowbiz/article-5874213/Did-Miley-Cyrus-Liam-Hemsworth-secretly-married.html		Did Miley Cyrus and Liam Hemsworth secretly get married?	0	0
https://s.yimg.com/uu/api/res/1.2/rNLbr_QGKbcHp3fAhfgMXw--~B/aD0xNjY5O3c9MjUwMDtzbT0xO2FwcGlkPXl0YWNoeW9u/https://media.zenfs.com/en-US/nbc_news_122/b4cdec47217b1409c2ab2bd31284fdae		text voter afraid of contract the coronavirus can cast their ballot by mail in upcom elect fewer judge rule tuesday in page decks you district judge fred bieri wrote that the coronavirus tandem had left the world without immune and fear disable the decks came after the text democrat parti and individu voter file suit last month argue that the state denial of mail in ballot sure tandem was violet of their constitute right only people over the age of and those with disable that prevent them from vote in person are allow to cast mail in ballot in the state echo fall claim made by preside	0	1

5. CONCLUSION

This study proposes an efficient model to detect text-image fake news by improving the extraction from text and image inputs. The texts are handled in a way that facilitates the extraction of their influential words, especially in cases where the text length exceeds 512 words, which exceeds the capacity of most text representation models, such as BERT and CLIP. Thus, we overcome the issue of cutting these texts and not benefiting from words that may be important but come after the maximum length allowed to handle. The system for determining the reality news from its falsity uses many DL techniques. It combines them in an efficient way that reflects positively on the features extracted from the text and image pairs entered into the proposed system. After including the texts and images inserted using CLIP embedding, we relied on passing the CLIP embedding of text parallel to the CNN model based on the Residual Block and the LSTM model. The extracted features result (CNN features, LSTM features, and CLIP embedding features) are passed to a dense layer (text dense layer). Also, the CLIP embedding of the image is passed to the dense layer (image dense layer) before sending the result features of it with the other dense result (text dense layer) to the cross-attention mechanism layer to understand the relationship between the image and text features

in parallel. The next stage involves using GFN by passing three inputs: cross-attention result, text CNN features, and text CLIP embedding features. Finally, we get the most influential features, which are transferred to a fully connected layer to determine whether the news is real. The model was tested on open-source English training sets RECOVERY, GOSSIPCOP, and Ti-CNN. The experimental results, in which we obtained accuracy values of 90 for the GOSSIPCOP dataset and 99 for the Ti-CNN dataset, indicate that the proposed model performed well compared to other classical models, demonstrating its ability to detect fake news in the future and trying to develop it after overcoming the most big obstacles it faced, the most important of which is the availability of a large dataset and modern data that has a diversity of languages and dialects as well as the news topics it contains. We also concluded that it can distinguish fake news whenever the input texts are shorter in length (Ti-CNN dataset, which is the shortest length text used here) due to its ability to summarize selected words in a length that is proportional to the maximum length allowed for the input text for the embedding mechanisms (like CLIP), which is 512 tokens. All these requirements enhance the proposed models' ability to detect fake news in live applications.

REFERENCES

[1] Yadav A, Gaba S, Khan H, Budhiraja I, Singh A, Singh KK (2024) ETMA: Efficient transformer-based multilevel attention framework for multimodal fake news detection. IEEE Trans Comput Soc Syst. 11(4):5015-5027. doi:10.1109/tcss.2023.3255242.

[2] Abduljaleel IQ, Ali IH (2024) Deep learning and fusion mechanism-based multimodal fake news detection methodologies: A review. Eng Technol Appl Sci Res. 14(4):15665-15675. doi:10.48084/etasr.7907.

[3] Nawaz MZ, Nawaz MS, Fournier-Viger P, He Y (2024) Analysis and classification of fake news using sequential pattern mining. Big Data Min Anal. 7(3):942-963. doi:10.26599/bdma.2024.9020015.

[4] Abbood EA, Al-Assadi TA (2022) A new convolution neural layer based on weights constraints. In: 2022 International Conference on Data Science and Intelligent Computing (ICDSIC). IEEE.

[5] Abduljaleel IQ, Ali, IH (2025) Detecting fake news using BERT word embedding, attention mechanism, partition and overlapping text techniques. TEM Journal, 1152–1165. https://doi.org/10.18421/tem142-16.

[6] Gu Y, Castro I, Tyson G (2024) Detecting Multimodal Fake News with Gated Variational AutoEncoder. In: ACM Web Science Conference. ACM; 129-138. doi: https://doi.org/10.1145/3614419.3643992.

[7] Lu J, Goswami V, Rohrbach M, Parikh D, Lee S (2020) 12-in-1: Multi-Task Vision and Language Representation Learning. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). doi: 10.1109/CVPR42600.2020.01045.

[8] Yan F, Zhang M, Wei B, Ren K, Jiang W (2024) FMC: Multimodal fake news detection based on multi-granularity feature fusion and contrastive learning. Alex Eng J.109:376-393. doi:10.1016/j.aej.2024.08.103.

[9] Kaliyar RK, Goswami A, Narang P (2021) FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. Multimed Tools Appl. ;80(8):11765-11788. doi:10.1007/s11042-020-10183-2.

[10] Rustam F, Aljedaani W, Jurcut AD, Alfarhood S, Safran M, Ashraf I (2024) Fake news detection using enhanced features through text to image transformation with customized models. Discov Computing.;27(1). doi:10.1007/s10791-024-09490-1.

[11] Dai K, Shao J, Gong B, Jing L, Chen Y (2024) CLIP-FSSC: A transferable visual model for fish and shrimp species classification based on natural language supervision. Aquacult Eng. ;107(102460):102460. doi:10.1016/j.aquaeng.2024.102460.

[12] Luvembe AM, Li W, Li S, Liu F, Wu X (2024) CAF-ODNN: Complementary attention fusion with optimized deep neural network for multimodal fake news detection. Inf Process Manag. 61(3):103653. doi:10.1016/j.ipm.2024.103653.

[13] Yang H, Zhang J, Zhang L, Cheng X, Hu Z (2024) MRAN: Multimodal relationship-aware attention network for fake news detection. Computer Stand Interfaces. 89(103822):103822. doi:10.1016/j.csi.2023.103822.

[14] Yan F, Zhang M, Wei B, Ren K, Jiang W (2024) SARD: Fake news detection based on CLIP contrastive learning and multimodal semantic alignment. J King Saud Univ - Computer Inf Sci. ;36(8):102160. doi:10.1016/j.jksuci.2024.102160.

[15] Qu Z, Meng Y, Muhammad G, Tiwari P (2024) QMFND: A quantum multimodal fusion-based fake news detection model for social media. Inf Fusion. 104(102172):102172. doi:10.1016/j.inffus.2023.102172.

[16] Jonnapalli TR, Selvi M (2024) Detecting fake news in social media networks with deep learning techniques. In: AIP Conference Proceedings. Vol 3086. AIP Publishing;030006. doi: https://doi.org/10.1063/5.0211567.

[17] Rashid J, Kim J, Masood A (2024) Unraveling the tangle of disinformation: A multimodal approach for fake news identification on social media. In: Companion Proceedings of the ACM Web Conference 2024. Vol 3. ACM; 1849-1853. doi: https://doi.org/10.1145/3589335.3651972.

[18] Bahdanau D, Cho K, Bengio Y (2015) Neural machine translation by jointly learning to align and translate. In: Proceedings of the 3rd International Conference on Learning Representations, pp. 1–15. doi: https://doi.org/10.48550/arXiv.1409.0473.

[19] Kou F, Wang B, Li H, Zhu C, Shi L, Zhang J, et al (2025) Potential Features Fusion Network for multimodal fake news detection. ACM Trans Multimedia Computer Communication Appl. doi:10.1145/3711866.

[20] Wei K, Dai J, Hong D, Ye Y (2024) MGFNet: An MLP-dominated gated fusion network for semantic segmentation of high-resolution multi-modal remote sensing images. Int J Appl Earth Obs Geoinf. 135(104241):104241. doi:10.1016/j.jag.2024.104241.

[21] Zhang C, Wu J (2024) Software defect prediction based on effective fusion of multiple features. IEEE Access. Published online 2024:1-1. doi:10.1109/access.2024.3409709.

[22] Singhal S, Shah RR, Chakraborty T, Kumaraguru P, and Satoh S (2019) Spotfake: A multi-modal framework for fake news detection. in: 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM), IEEE, pp. 39–47. Doi: https://doi.org/10.1109/BigMM.2019.00-44.

[23] Silva A, Luo L, Karunasekera S, and Leckie C (2021) Embracing domain differences in fake news: Cross-domain fake news detection using multi-modal data. Proc. Conf. AAAI Artificial Intelligence., vol. 35, no. 1, pp. 557–565. doi: https://doi.org/10.1609/aaai.v35i1.16134.

[24] Chen Y, Li D, Zhang P, Sui J, Lv Q, Tun L, and Shang L (2022) Cross-modal ambiguity learning for multimodal fake news detection. in: Proceedings of the ACM Web Conference 2022, pp. 2897–2905. Doi: https://doi.org/10.1145/3485447.3511968.

https://doi.org/10.31436/iiumej.v26i3.3629

DEVELOPMENT OF A PREDICTIVE REAL-TIME TEMPERATURE MANAGEMENT FRAMEWORK FOR DATA CENTERS USING A HYBRID DEEP LEARNING MODEL

PATRICK O. OLABISI*, AHMED A. MAYOWA

Department of Electrical, Electronic and Telecommunication Engineering, College of Engineering, Bells University of Technology, Ota, Ogun State, Nigeria

**Corresponding author: niyiolabisi@gmail.com*

(Received: 28 February 2025; Accepted: 22 July 2025; Published online: 9 September 2025)

ABSTRACT: Traditional cooling systems in data centers often struggle to adapt to dynamic temperature fluctuations caused by varying server loads and environmental conditions, leading to energy inefficiency and potential overheating risks. This research developed a predictive cooling system to address the challenges of maintaining efficient temperature management that will enhance system reliability, cooling efficiency, and energy savings by leveraging deep learning techniques. A hybridization of two deep learning techniques: Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks was formed. Over a 24-hour test period, the model demonstrated a high prediction accuracy ranging from 85 to 95% with reliability between 0.97 and 0.99, highlighting the system's robust design. Cooling efficiency ranged from 0.75 to 0.9, peaking during cooler hours and dipping slightly during midday due to higher energy demands and environmental conditions. Efficiency improvements of 3 to 7% were observed, particularly during high-demand periods. The findings highlight the transformative potential of predictive cooling systems in reducing energy consumption and ensuring thermal stability in modern data centers. Further optimization of the CNN-LSTM model to improve prediction accuracy during peak server activity is recommended. Additionally, integrating real-time feedback loops and incorporating renewable energy sources into the system could enhance its sustainability and reduce operational costs.

ABSTRAK: Sistem penyejukan tradisional di pusat data sering menghadapi masalah menyesuaikan diri dengan suhu dinamik yang disebabkan oleh beban pelayan dan keadaan persekitaran berbeza, membawa kepada risiko ketidakcekapan tenaga dan kemungkinan terlebih panas. Penyelidikan ini membangunkan sistem penyejukan ramalan bagi menangani cabaran dalam mengekalkan pengurusan suhu yang cekap, sekaligus meningkatkan kebolehpercayaan pada sistem, kecekapan penyejukan dan penjimatan tenaga melalui manfaat teknik pembelajaran mendalam. Hibridisasi dua teknik pembelajaran mendalam iaitu: Rangkaian Neural Konvolusi (CNN) dan Rangkaian Memori Jangka Pendek Panjang (LSTM) telah dibentuk. Sepanjang tempoh ujian 24 jam, model menunjukkan ketepatan ramalan yang tinggi antara 85 hingga 95% dengan kebolehpercayaan antara 0.97 dan 0.99 menyerlahkan reka bentuk sistem yang teguh. Kecekapan penyejukan antara 0.75 hingga 0.9, memuncak pada waktu sejuk dan menurun sedikit pada tengah hari disebabkan perningkatan permintaan tenaga dan keadaan persekitaran. Peningkatan kecekapan sebanyak 3 hingga 7% diperhatikan, terutama semasa tempoh permintaan tinggi. Dapatan ini menyerlahkan potensi transformatif sistem penyejukan ramalan dalam mengurangkan penggunaan tenaga dan memastikan kestabilan haba di pusat data moden. Pengoptimuman lanjut terhadap model CNN-LSTM bagi meningkatkan ketepatan ramalan semasa aktiviti waktu puncak adalah disyorkan. Selain itu, penyepaduan gelung maklum balas masa nyata dan gabungan sumber

tenaga boleh diperbaharui ke dalam sistem bagi meningkatkan lagi kemampanannya dan mengurangkan kos operasi.

KEYWORDS: *CNN, LSTM, Temperature Management, Cooling Systems, Data Centers*

1. INTRODUCTION

Temperature fluctuations in data centers often follow specific trends based on server usage patterns, peak hours of activity, and external environmental factors such as ambient temperature. Cooling technologies in data centers have gone from dry air cooling to liquid cooling, immersion, evaporative cooling, and so on. While traditional air-based systems remain prevalent, liquid-based, hybrid, and advanced cooling methods significantly improve efficiency and performance. Despite the advancements in cooling technologies in data centers, several limitations remain, particularly when balancing cost, complexity, efficiency, and environmental constraints. Traditional air cooling systems, for example, struggle to keep up with the increasing power density of modern data centers. As servers become more powerful, they generate more heat, and air cooling systems often become inefficient in effectively removing the heat. This inefficiency leads to higher energy consumption, making it difficult for data centers to reduce operational costs and meet sustainability goals [1].

Moreover, as the demand for high-performance computing and data storage continues to rise and data centers strive to become more energy-efficient, the growing need for cooling systems that not only manage heat effectively but also minimize their own energy consumption and are scalable and environmentally friendly has become more indispensable [2]. The challenge lies in developing cooling technologies that balance performance, cost, and sustainability.

As effective as they are, traditional cooling methods often operate on static principles, are reactive, and only respond to temperature changes after they have occurred. This approach can result in inefficiencies, as systems either overcompensate by cooling more than necessary or lag in responding to rising temperatures, putting equipment at risk [3]. Artificial Intelligence (AI) approach offers a more dynamic, proactive, and adaptive approach to temperature management, using predictive analytics and real-time decision-making to ensure that data centers remain energy-efficient and operationally resilient [4]. Temperature management is a dynamic process, with environmental conditions fluctuating rapidly based on server activity, external factors, and equipment performance. AI systems can continuously monitor these conditions and make real-time adjustments to maintain optimal temperatures while minimizing energy consumption.

The use of artificial intelligence (AI) has, in recent years, made a significant impact on data center management, particularly in temperature regulation. The complexity of maintaining optimal temperatures in a data center, where hundreds or even thousands of servers generate massive amounts of heat, has led to a growing reliance on intelligent systems that can make real-time predictions and adjustments. AI has found numerous applications in data center operations, from automating mere tasks to optimizing complex processes. In temperature management, AI techniques are employed to monitor, analyze, and control environmental conditions. Machine learning (ML) algorithms, deep learning models, and neural networks are some of the most commonly used AI techniques and intelligent computational tools that are used to optimize the cooling systems in data centers [5], [6].

A significant advantage of AI is its ability to process vast amounts of data and identify patterns that are otherwise difficult for humans or traditional systems to detect. Data centers

produce extensive data streams, including information on temperature fluctuations, energy consumption, airflow, and server workload. AI-powered systems can analyze these data streams in real time and adjust cooling systems accordingly, optimizing the balance between energy consumption and maintaining the necessary environmental conditions. AI algorithms can predict future temperature trends based on historical data and current operating conditions. This allows data centers to proactively adjust their cooling systems to prevent overheating or inefficiencies in energy consumption before they occur. By anticipating temperature changes, AI reduces the likelihood of thermal spikes leading to equipment failure or energy wastage. Predictive maintenance, driven by AI, helps forecast when cooling system repair or adjustment is necessary, thus avoiding unnecessary downtime, reducing maintenance costs, and minimizing energy waste [7]. Predictive analytics also helps optimize server workloads by forecasting which equipment will most likely heat up and distributing tasks accordingly, ensuring no single server is overburdened [3]. This dynamic allocation of workloads ensures that servers operate within their optimal temperature ranges and contributes to energy efficiency across the facility.

AI-powered cooling systems use feedback loops to respond instantaneously to temperature or server load changes. For example, suppose a specific rack of servers begins to overheat due to increased computational demands. In that case, the AI system can immediately adjust the airflow or cooling distribution to target that area without unnecessarily cooling the entire facility. This targeted cooling approach reduces the overall energy consumption of the data center, as it avoids the wasteful practice of cooling areas that do not require it [8].

2. DEEP LEARNING IN ENVIRONMENTAL CONTROL SYSTEMS

Deep learning (DL), as a subset of machine learning, is characterized by its ability to learn from vast amounts of data through layered architectures of artificial neural networks. The networks consist of an input layer, multiple hidden layers, and an output layer interconnected by nodes or neurons that transform input data by applying weights and activation functions, as shown in Figure 1 [9]. The most common type of deep learning model is the feed-forward neural network, where data moves in one direction – from input to output. However, several specialized deep learning models have performed exceptionally well in specific tasks.

Deep learning techniques are classified into supervised (discriminative), unsupervised (generative), and hybrid techniques as shown in Figure 2 [10]. Major techniques in the discriminative category include Multi-Layer Perceptron (MLP), Convolutional Neural Networks (CNN or ConvNet), and Recurrent Neural Networks (RNN), with the variants of RNN being Long Short-Term Memory (LSTM), Bidirectional RNN/LSTM (Bi-LSTM), and Gated Recurrent Units (GRUs). The unsupervised techniques, which include Generative Adversarial Network (GAN), Autoencoder (AE), Restricted Boltzmann Machine (RBM), Self-Organizing Map (SOM), and Deep Belief Network (DBN), and their variants, are usually used for high-order correlation of features of patterns and joint statistical distributions of data and their classes. The variants of the Autoencoder technique are the Variational Autoencoder (VAE), Sparse Autoencoder (SAE), Denoising Autoencoder (DAE), Contractive Autoencoder (CAE-t), Undercomplete Autoencoder, Convolutional Autoencoder (CAE-v), and Adversarial Autoencoder (AAE). The hybrid deep learning techniques include a hybrid of deep learning techniques and several others, like deep transfer learning (DTL) and deep reinforcement learning (DRL).

https://doi.org/10.31436/iiumej.v26i3.3629

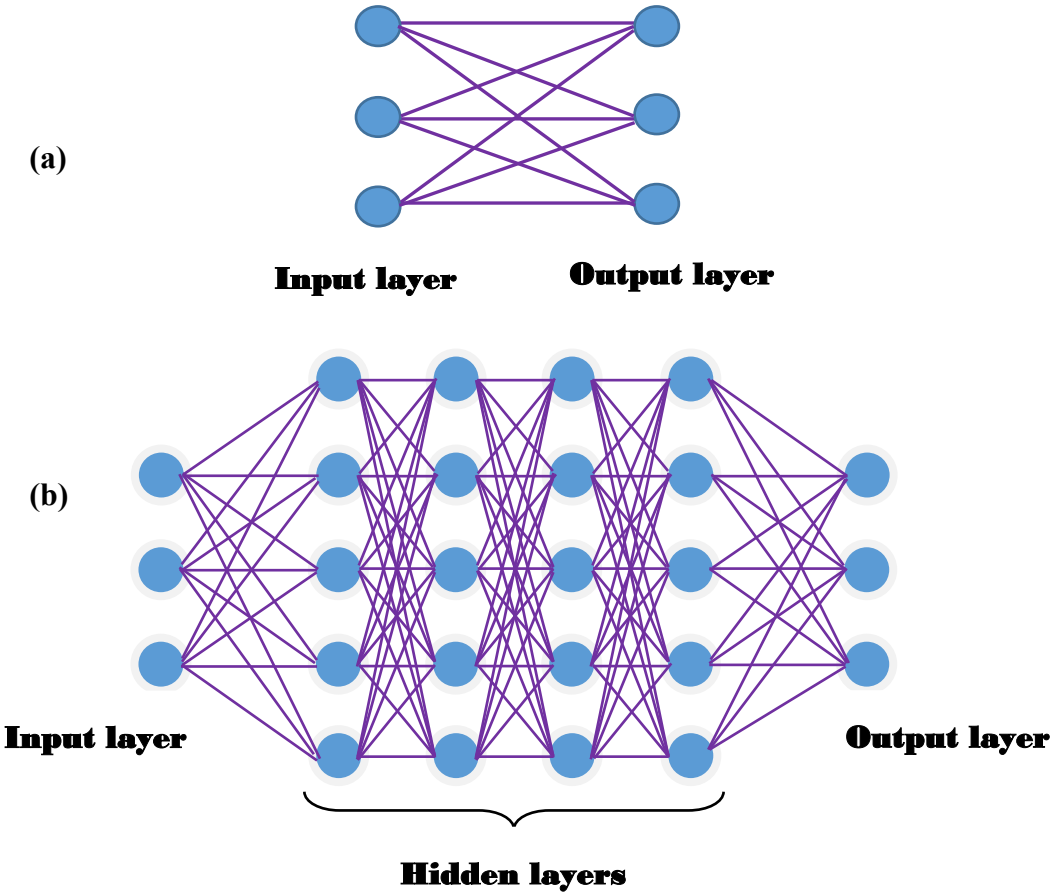


Figure 1. Neural networks (a) Normal neural network (b) Deep learning neural network [9].

Deep learning has emerged as a powerful tool for analyzing and managing the complex dynamics of environmental control systems, particularly in data centers. With modern data centers' increasing scale and complexity, traditional environmental management methods, such as manual monitoring and static control systems, have become less efficient. Deep learning models offer a highly sophisticated solution for optimizing temperature regulation and other environmental parameters [11].

Unlike conventional machine learning techniques, using multi-layered neural networks in DL allows the models to learn abstract representations from vast amounts of data. This makes DL particularly useful in environments like data centers, where temperature regulation involves numerous interacting factors, such as server workload, cooling system performance, power consumption, and external conditions. By employing deep learning models, data centers can achieve more precise and adaptive control over their environmental systems, resulting in improved energy efficiency, reduced cooling costs, and extended equipment life. Deep learning models offer an advanced approach to managing this complexity, providing highly accurate predictions and enabling real-time optimization of cooling strategies [12].

One of the key advantages of deep learning models in temperature regulation is their ability to learn and adapt to changing conditions continuously. As new data is collected from temperature sensors, server logs, and cooling system monitors, the models can use the data to update their predictions and recommendations in real time [13]. This adaptive capability is significant in modern data centers, where server workloads and environmental conditions can change rapidly. By continuously adjusting to these changes, deep learning models ensure that data centers maintain optimal environmental conditions while minimizing energy consumption

https://doi.org/10.31436/iiumej.v26i3.3629

and cooling costs. In addition, DL models are highly scalable and can be applied to data centers of any size or complexity in handling the vast amounts of data generated by large data centers, making them an ideal solution for managing environmental control in such environments [14].

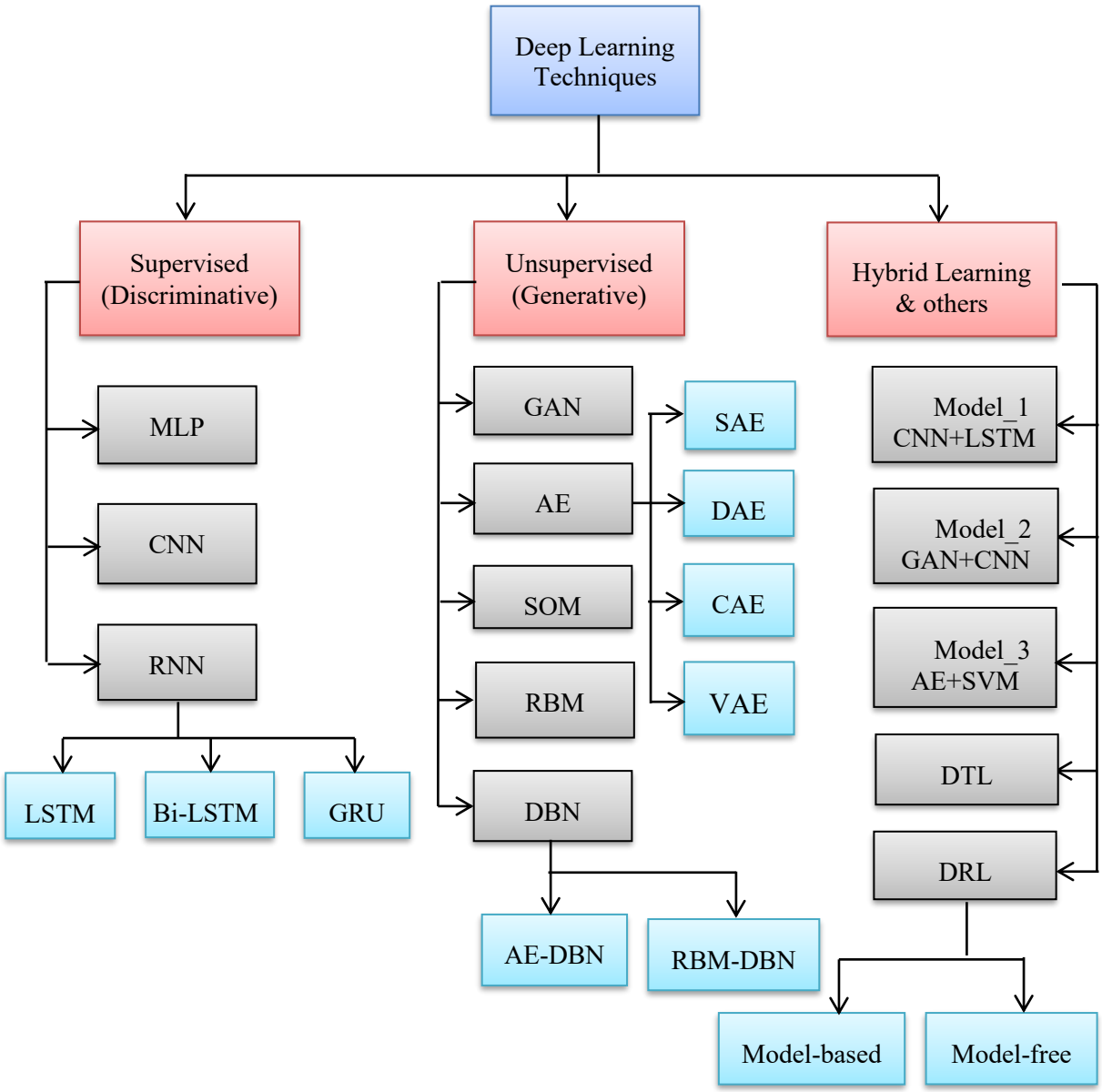


Figure 2. Deep Learning Techniques according to the Categories [10].

Therefore, deep learning models' ability to process large datasets, learn intricate patterns, and generalize from data makes them well-suited to provide efficient and powerful solutions for managing the complex data patterns involved in temperature regulation and the dynamic and data-rich environment of modern data centers. With these models, data centers can achieve more precise and adaptive control over their environmental systems, improving energy efficiency, reducing cooling costs, and prolonging the lifespan of critical IT infrastructure.

Supervised DL techniques of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks stand out for their ability to handle large, complex datasets, making them ideal for modern data centers' dynamic and data-rich environment. CNNs are

particularly effective in tasks that involve spatial data, making them well-suited for analyzing temperature distributions across different areas of a data center [11]. The convolutional operation is mathematically given by [9]:

$$f_l^k(p,q) = \sum_c \sum_{x,y} i_c\,(x,y) \odot e_l^k(u,v) \quad (1)$$

CNNs can identify areas at risk of overheating, recommend targeted cooling strategies, and capture spatial dependencies such as how airflow and cooling efficiency vary across racks or server units, which is crucial for optimizing localized cooling strategies. As shown in Figure 3 [15], CNNs' layered architecture allows them to learn intricate patterns in the data that would be difficult for traditional models to capture, making them indispensable tools in the optimization of environmental control systems.

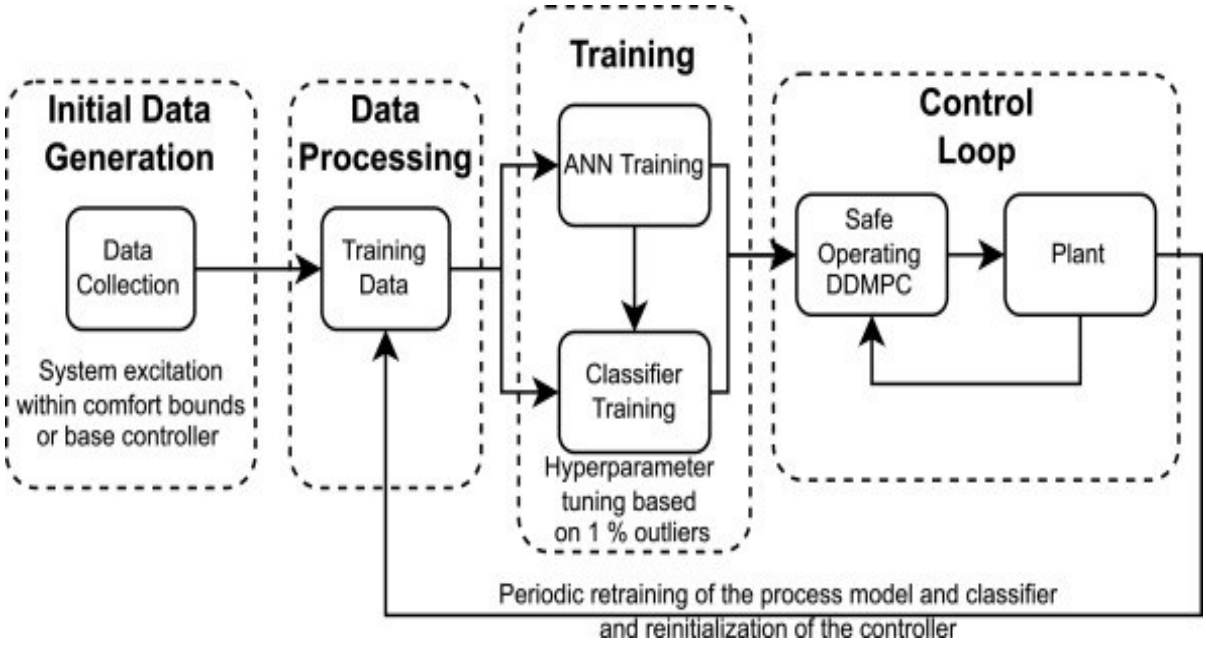


Figure 3. CNN Data Center Temperature Control [15].

On the other hand, the LSTM networks shown in Figure 4 [16], as a variant of a recurrent neural network (RNN), are specifically designed to handle long-term dependencies in sequential data and so are ideal for tasks that involve time series data, such as temperature prediction over time. This makes them well-suited for temperature regulation tasks where past temperature trends, server workloads, and cooling system performance influence future temperature outcomes. Hence, the LSTM networks can capture temporal dependencies in time series data, allowing them to predict temperature trends based on past observations and anticipate future fluctuations [17], [18]. By utilizing LSTM models, data centers can achieve highly accurate forecasts of temperature fluctuations, allowing for proactive adjustments to cooling systems before critical temperature thresholds are reached [19]. LSTM models can also simulate the impact of different cooling strategies, helping operators identify the most energy-efficient and cost-effective solutions.

The hybrid architecture of CNN-LSTM leverages the powerful feature extraction capabilities of convolutional layers and the ability of LSTM to capture sequential dependencies. The CNN component processes input data, such as temperature, server workload, and cooling patterns, extracting critical features. These features are then passed to the LSTM, which analyzes temporal trends and predicts future temperature fluctuations [8].

https://doi.org/10.31436/iiumej.v26i3.3629

Figure 5 shows the flow chart of the operations of the hybrid model [20]. Based on these predictions, the model adjusts cooling systems in real time, optimizing energy usage and preventing overheating. This proactive control ensures consistent cooling efficiency, minimizes energy consumption, and maintains operational stability, effectively adapting to the data center's dynamic environmental and server load conditions.

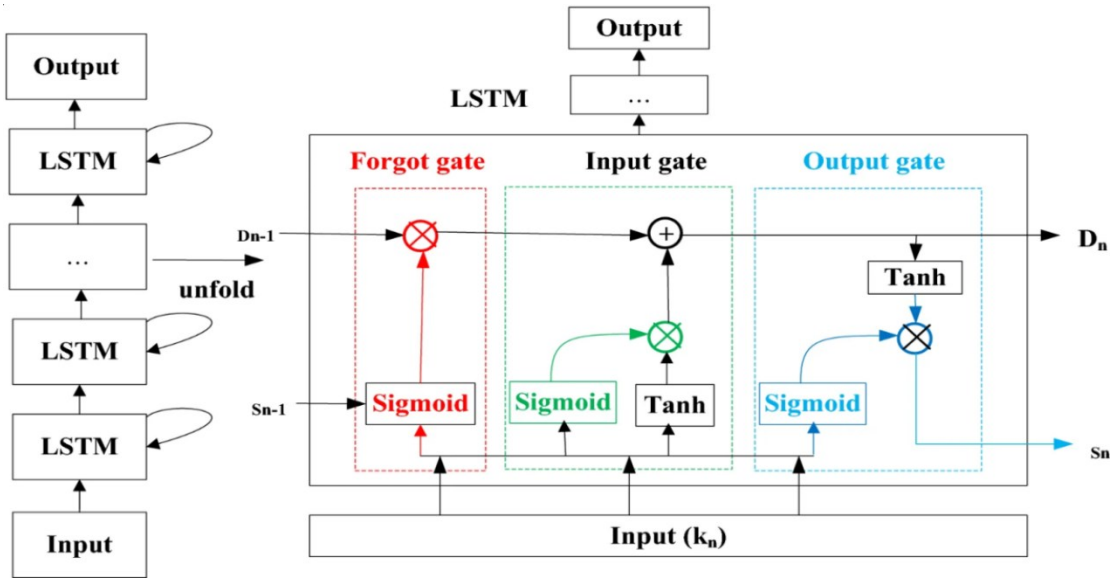


Figure 4. Cell logic structure of the LSTM network [16].

The CNN-LSTM hybrid model is proposed for this work as it has been demonstrated to outperform other models in other heat-related studies by different researchers. From the review of literature done for this work, literature on the use of a CNN-LSTM hybrid model for temperature predictions in data centers is scarce. In the work of Elmaz *et al.* [21], CNN-LSTM outperformed other models used in optimizing Heating, Ventilation, and Air Conditioning (HVAC) systems across all time frames and demonstrated greater resilience to error accumulation, achieving an accuracy of 0.9 for the 120-minute prediction horizon. Guo *et al.* [16] simulated climate parameters for Jinan city in China and forecasted six climatic factors monthly using data over 72 years to study climate change impacts on various aspects of life using several DL models. The authors found that the CNN-LSTM hybrid model achieves higher accuracy and significantly reduces forecasting errors than the other models of artificial neural networks (ANN), recurrent neural networks (RNN), long short-term memory neural networks (LSTM), and deep convolutional neural networks (CNN) that they used. The findings demonstrate the potential of CNN-LSTM models to enhance climate forecasting, which can aid in meteorological disaster prevention and reduction, as well as flood control and drought mitigation. Figure 6 shows the working architecture of the hybrid CNN-LSTM model [22].

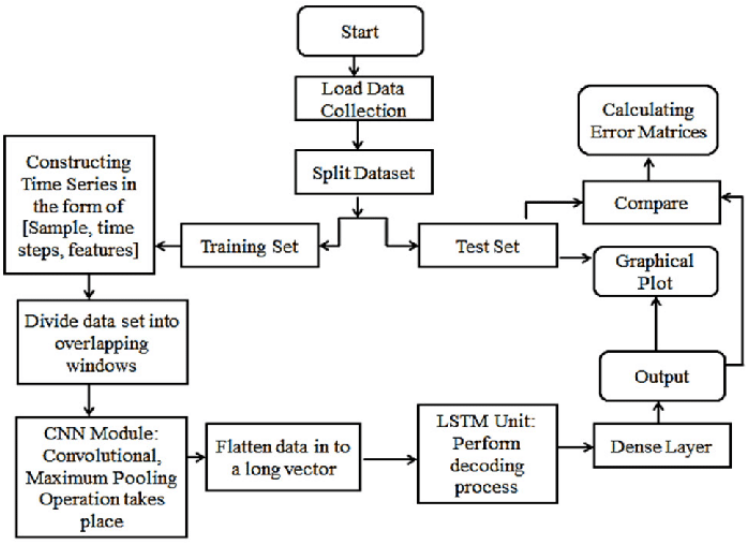


Figure 5. CNN-LSTM model data selection, feature extraction, and decoding process [20].

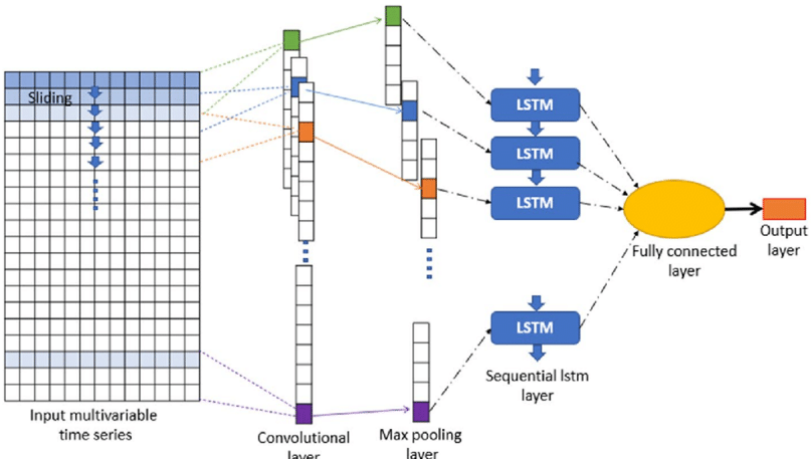


Figure 6. Architecture of the CNN-LSTM Model [22].

3. METHODS

3.1. Design of a Predictive Temperature Management System using CNN-LSTM

The output of the convolutional layer $O_{CNN}(t)$ in the CNN-LSTM model is the convolution of the input data I(t) with a filter F, given by:

$$O_{CNN}(t) = \ I(t) * F \tag{2}$$

where, $O_{CNN}(t)$ is the CNN output layer (no unit), $I(t)$ is the input temperature data (no unit), F is convolution filter (no unit).

The LSTM model predicts temperature $T_{LSTM}(t)$ by utilizing previous outputs and cell states $C(t-1)$ for time-series data processing and it is given by:

$$T_{LSTM}(t) = f(C(t-1), h(t-1)) \tag{3}$$

where, $T_{LSTM}(t)$ is the predicted temperature (in °C), $C(t-1)$ is cell state from previous time-step (no unit) and $h(t-1)$ is the hidden state from previous time-step (no unit).

https://doi.org/10.31436/iiumej.v26i3.3629

The loss function value L for training the CNN-LSTM model is the mean squared error (MSE) between the predicted $T_{LSTM}(t)$ and actual temperature values $T_{actual}(t)^2$. It is given by:

$$L = \frac{1}{N}\sum_{i=1}^{N}(T_{LSTM}(t) - T_{actual}(t))^2 \quad (4)$$

where, $T_{actual}(t)^2$ is actual temperature at time t (in °C).

The parameter of the CNN-LSTM model based on θ is updated using gradient descent as function of learning rate η and the gradient of the loss function [23], is given as Equation (5):

$$\theta_{new} = \theta_{old} - \eta\frac{\partial L}{\partial \theta} \quad (5)$$

where, θ is model parameters, η is Learning rate, and L is Loss function (all have no units).

Temperature input data $T_{norm}(t)$ is normalized for training the CNN-LSTM model using the mean μ and standard deviation σ and given by [24]:

$$T_{norm}(t) = \frac{T(t)-\mu}{\sigma} \quad (6)$$

where, $T_{norm}(t)$ is normalized temperature (no unit), $T(t)$ is actual temperature at time t (in °C), μ is mean temperature (in °C) and σ is standard deviation of temperature (in °C).

3.2. Implementation and integration of the CNN-LSTM model for real-time temperature prediction in data centers.

In this sub-section, the CNN-LSTM model is implemented for real-time temperature prediction in data centers and integrated into the existing cooling system. The predicted temperature $T_{pred}(t)$ at any time t (in °C), is computed by feeding real-time input data into the trained CNN-LSTM model and given by:

$$T_{pred}(t) = CNN_LSTM(I(t)) \quad (7)$$

where, $I(t)$ is the real-time input data (no unit).

The cooling system activation function, $A(t)$, depends on the predicted temperature $T_{pred}(t)$ crossing a threshold T_{th}, and is given by:

$$A(t) = H(T_{pred}(t) - T_{th}) \quad (8)$$

where, H is the Heaviside step function, T_{th} is the temperature threshold for activation (in °C).

The total energy consumption at time t, $E_{pred}(t)$ (in kWh), with predictive cooling is the sum of server energy usage E_s and cooling energy E_c modulated by predicted cooling cycles, is given by Equation (13):

$$E_{pred}(t) = E_s(t) + A(t) \times E_c(t) \quad (9)$$

where, $E_s(t)$ is server energy consumption at time t (in kWh) and $E_c(t)$ is cooling energy consumption (in kWh).

The response time t_r (in seconds), is the delay between the prediction and cooling system activation, dependent on system latency L_s and cooling efficiency η_c, is given by:

$$t_r = \frac{L_s}{\eta_c} \quad (10)$$

where, L_s is system latency (in seconds) and η_c cooling efficiency (no unit).

https://doi.org/10.31436/iiumej.v26i3.3629

The feedback temperature $T_{fb}(t)$ (in °C), is the measured real-time temperature used to correct the model's predictions, with ϵ representing the error term, and is given by:

$$T_{fb}(t) = T(t) + \epsilon(t) \tag{11}$$

where, $T(t)$ is measured temperature at time t (in °C) and ϵ(t) is prediction error (in °C).

The overheating probability, P_o is based on the temperature exceeding the critical threshold T_{crit} for a given duration Δt.

$$P_o = \Pr\left(T_{pred}(t) > T_{crit}\right) \tag{12}$$

where, T_{pred} is predicted overheating temperature (in °C) and T_{crit} is the critical overheating temperature (in °C).

The hardware failure rate, $F(t)$ (in failures per hour), increases exponentially with temperature $T(t)$ according to Arrhenius' law, where k is a proportional constant, is given by

$$F(t) = F_0 \times e^{k/T(t)} \tag{13}$$

where, F_0 is Baseline failure rate (in failures per hour) and k is Proportional constant (no units).

Figure 7 shows the flowchart for implementing the hybrid CNN-LSTM model for data centers. The flowchart begins with data collection, where real-time environmental parameters such as temperature, humidity, and energy consumption are gathered through sensors. This data then undergoes preprocessing, cleaning, and normalization, and is suitable for machine learning models, ensuring accuracy and consistency.

The performance monitoring stage compares the real-time conditions with the model's predictions to ensure accuracy and energy efficiency, with the system constantly refining its decisions. The process completes and loops, providing continuous monitoring and real-time temperature regulation, preventing overheating, and improving data center energy efficiency, as shown in the flowchart. Figure 8 shows a diagram of the CNN-LSTM system's Simulink connection, while Figure 9 is a screenshot of the MATLAB interface of the code-based simulations of the CNN-LSTM data center optimization and temperature control.

3.3. Analysis of reduced energy consumption and cooling costs in data centers

The cooling optimization O_c (no unit), is proportional to the predicted energy savings S and the reduced hardware stress H_s over time t.

$$O_c = S \times H_s(t) \tag{14}$$

where, S is the energy savings (in kWh) and $H_s(t)$ is the reduced hardware stress (in kWh).

The total cooling cost reduction C_r (in currency), is the difference between the initial cost C_i and the AI-driven optimized cost C_o.

$$C_r = C_i - C_o \tag{15}$$

where, C_i is the initial cooling cost (in currency) and C_o is the optimized cooling cost (in currency).

The lifespan extension L_e of data center equipment is inversely proportional to the cumulative overheating probability P_o and hardware failure rate $F(t)$, is given by:

$$L_e = \frac{1}{P_o - F(t)} \tag{16}$$

https://doi.org/10.31436/iiumej.v26i3.3629

where, P_o is the overheating probability (no unit) and $F(t)$ is the failure rate (in failures per hour).

The operational efficiency O_e (no unit), increases with the ratio of predictive model accuracy A_{pred} to the energy savings S.

$$O_e = \frac{A_{pred}}{S} \tag{17}$$

where A_{pred} is the prediction accuracy (no unit), and S is the energy savings (in kWh).

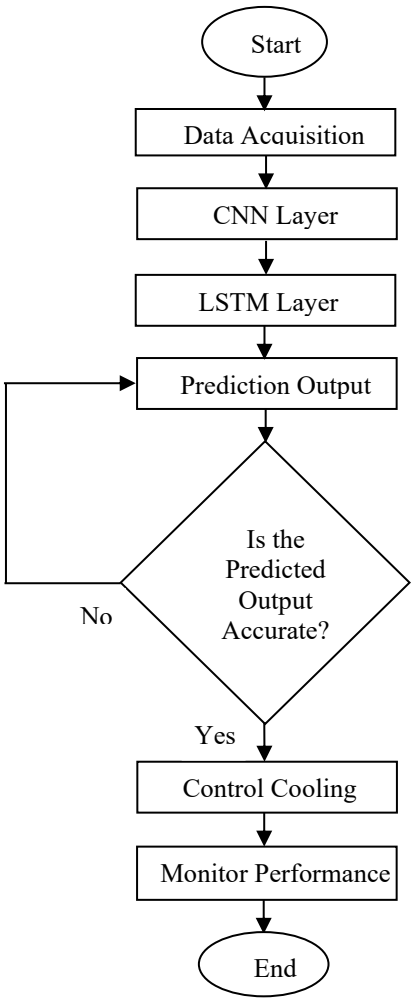


Figure 7. The Flowchart of the CNN-LSTM Temperature Control System.

https://doi.org/10.31436/iiumej.v26i3.3629

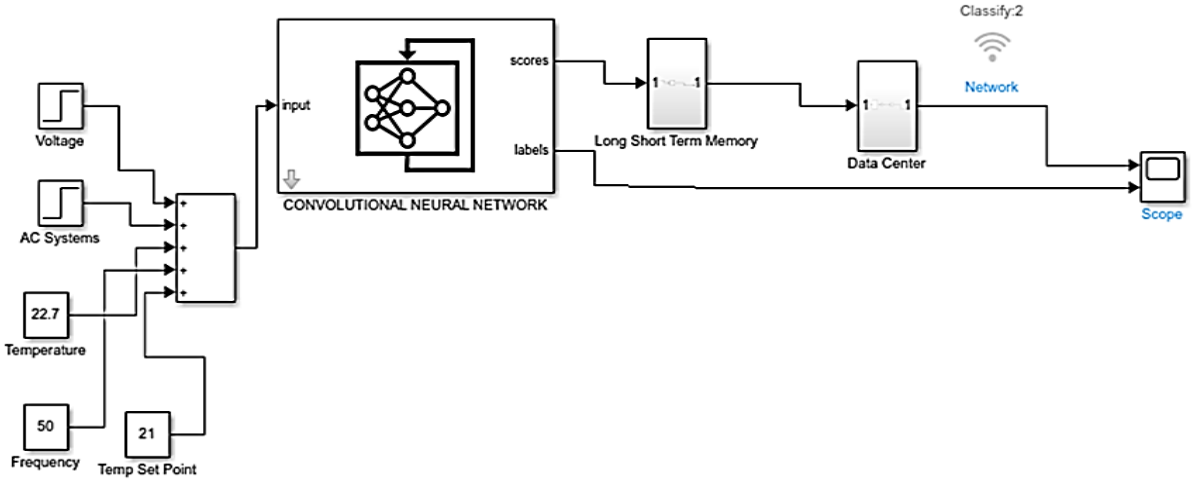


Figure 8. Simulink diagram of the CNN-LSTM System.

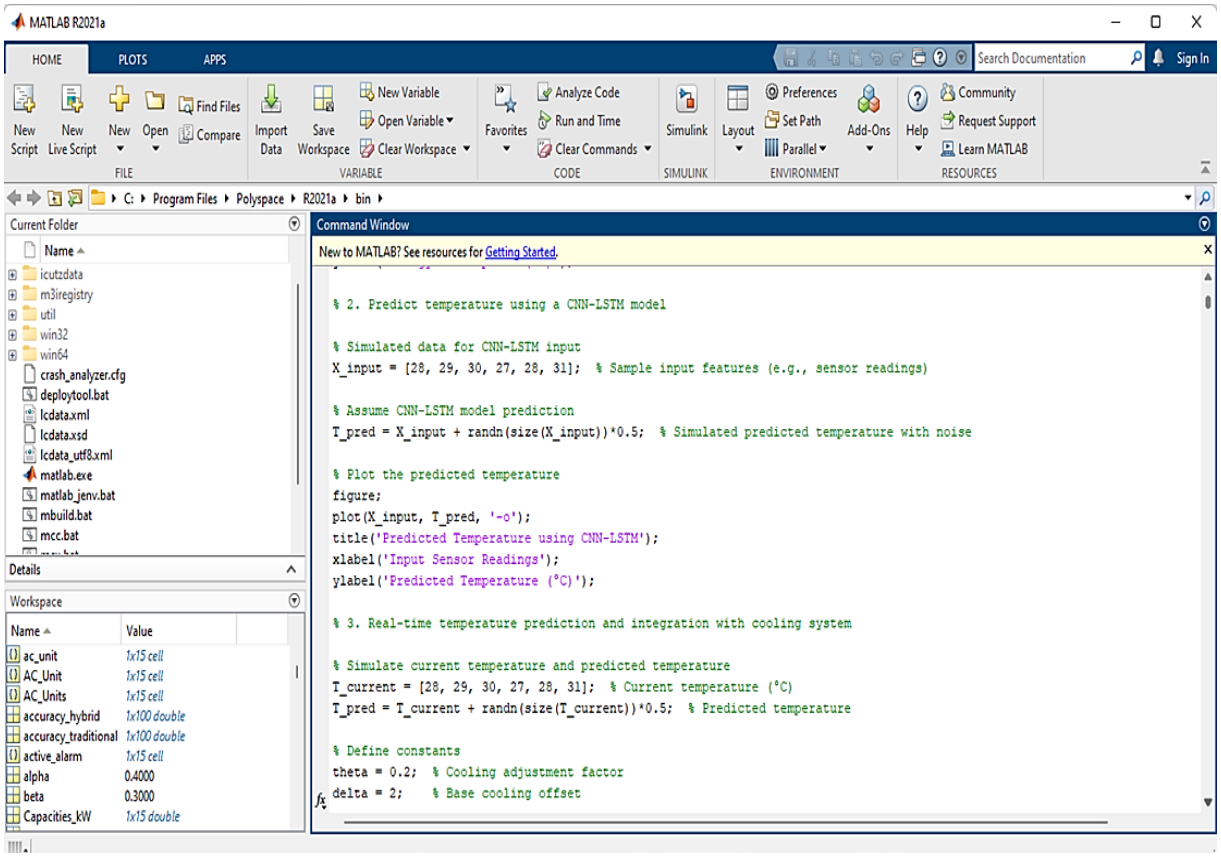


Figure 9. MATLAB Code-base Simulation Interface.

The overall system reliability R_s is enhanced by reducing the hardware failure rate F(t) and lowering overheating probability P_o.

$$R_s = \frac{1}{F(t) \times P_o} \tag{18}$$

where, R_s is the system reliability (no unit), $F(t)$ is the failure rate (in failures per hour) and P_o is the overheating probability (no unit).

4. RESULTS AND DISCUSSIONS

4.1. Performance Evaluation of the Hybrid CNN-LSTM Model

4.1.1. Classification Metrics

The classification metrics highlight the predictive model's performance in correctly identifying labels in Figure 10. The model's accuracy is obtained as 0.9, meaning the model correctly predicted 90% of the labels, showcasing its reliability. The precision, which measures the proportion of true positive predictions among all positive predictions, is 1.0. This indicates that the model made no false positive predictions, an ideal outcome for scenarios where precision is critical. The recall, or sensitivity, is calculated as 0.83, signifying that the model successfully identified 83% of the actual positive labels. While this value is slightly lower than precision, it demonstrates a solid ability to capture true positives. The F1-Score, a harmonic mean of precision and recall, is 0.91, balancing the trade-off between these two metrics and indicating an overall strong performance. The bar plot of these metrics, shown in Figure 10, visually emphasizes the model's strengths, with high accuracy, precision, and F1-score. At the same time, the slightly lower recall suggests room for improvement in capturing all positive labels. These quantitative results confirm the model's effectiveness, making it suitable for applications where precision and overall reliability are essential, though further tuning could enhance recall.

4.1.2. Error Metrics

The results shown in Figure 11 demonstrate the accuracy and reliability of the predictive model compared to the actual data, with error metrics providing a quantitative evaluation. The Root Mean Squared Error (RMSE), calculated as approximately 0.1, indicates a small average deviation between the predicted and actual values, reflecting the model's precision. Similarly, the Mean Absolute Error (MAE) is approximately 0.09, suggesting that, on average, the predictions deviate from the true values by less than 0.1 units. The Mean Squared Error (MSE), measuring the average squared differences, is approximately 0.008, highlighting the model's ability to minimize large errors effectively. The Coefficient of Determination (R^2), calculated as 0.995, showcases an excellent fit, with 99.5% of the variation in the actual data explained by the predictive model. This high value demonstrates the robustness of the model in capturing underlying patterns in the data. The bar plot of these metrics visually emphasizes the model's performance, where low RMSE, MAE, and MSE values align with the high R^2, providing confidence in the model's predictions. These results suggest that the predictive model achieves high accuracy, making it suitable for practical applications requiring reliable forecasting.

4.1.3. CNN-LSTM vs. Traditional Predictions: Quantitative Comparison

The plot shown in Figure 12 illustrates the accuracy of predictions made by CNN-LSTM and traditional methods against true sinusoidal values over 100 time units. For CNN-LSTM predictions, the Root Mean Square Error (RMSE) is 0.1201. This is significantly lower than the traditional method's value of 0.1892, indicating better precision. The Mean Absolute Error (MAE) for CNN-LSTM is 0.0956, compared to 0.1527 for the conventional method. The Mean Squared Error (MSE) is 0.0144 for CNN-LSTM predictions, reflecting reduced error magnitude relative to the conventional method's 0.0358. The R^2 value, a measure of model fit, is 0.9834 for the DL hybrid model, showing high accuracy, while the conventional method achieved a lower 0.9392. These values are presented in Table 1.

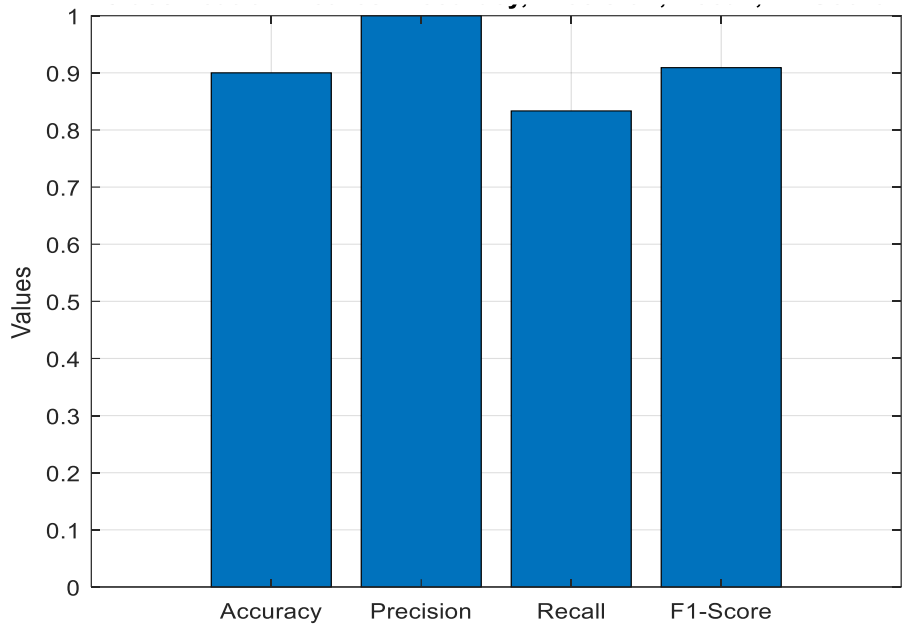


Figure 10. Classification Metrics: Accuracy, Precision, Recall, F1-Score.

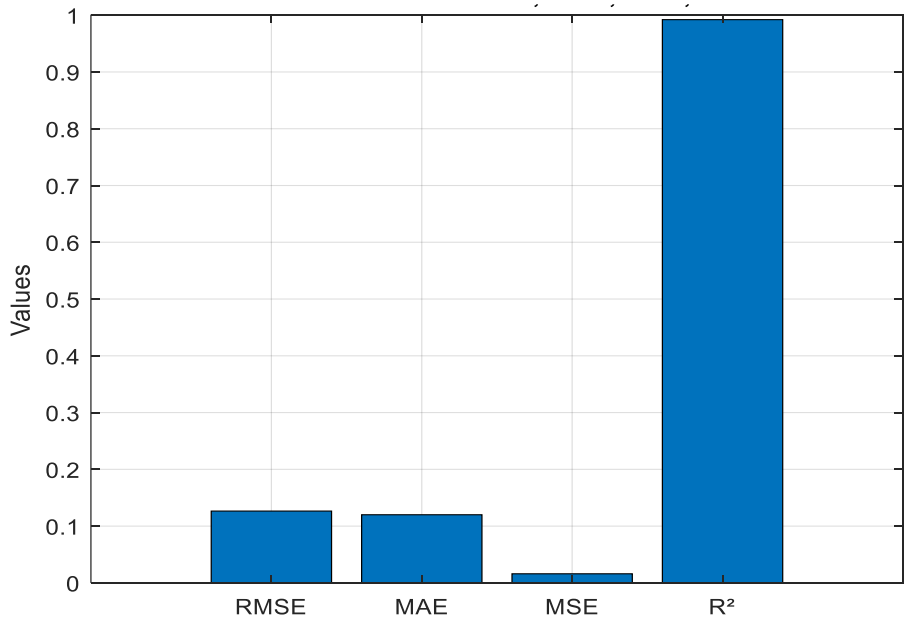


Figure 11. Error Metrics: RMSE, MAE, MSE, R^2.

4.2. CNN-LSTM vs. Traditional Predictions: Binary Classification Results

The binary prediction results compare the CNN-LSTM and traditional methods against 100 true labels (0s and 1s). The DL model achieved an accuracy of 0.90, outperforming the conventional method's result of 0.83, indicating a higher proportion of correct predictions. The precision of the DL predictions is 0.89, highlighting a better ability to identify positive cases, compared to the traditional method's 0.84. The DL's recall stands at 0.92, reflecting a strong ability to identify all true positives, surpassing the conventional method's 0.80. The DL-driven F1 score, which balances precision and recall, is 0.91, compared to the conventional method's 0.82, affirming CNN-LSTM's superior performance. Visual plots reveal CNN-LSTM predictions aligning closely with true labels, while traditional predictions show more deviation. With fewer false positives and negatives, the DL approach provides a robust solution for binary

classification, outperforming the conventional method across all key metrics as shown in Figure 13.

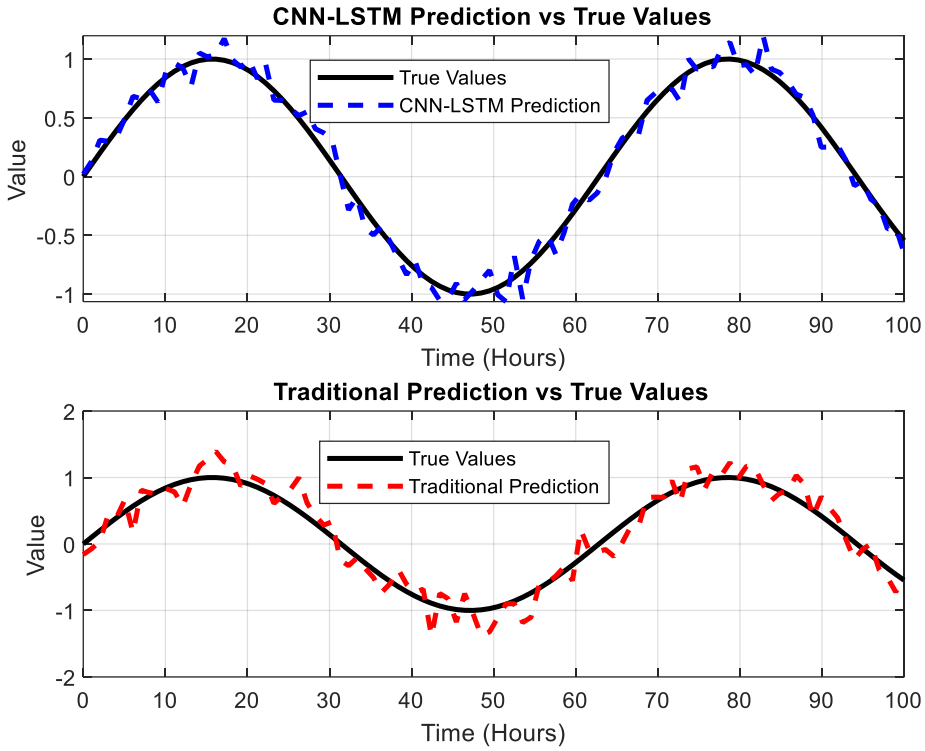


Figure 12. CNN-LSTM Prediction.

Table 1. Error Metric Evaluation

Metric	CNN-LSTM Prediction	Traditional Method
RMSE	0.098	0.1905
MAE	0.0762	0.1562
MSE	0.0096	0.0363
R^2	0.9783	0.9179

4.2. Analysis of the Data Centers' Energy Consumption and Cooling Costs

4.2.1. Temperature Reduction for AC1 to AC10

The results in Figure 14 model the temperature reduction for 10 air conditioning units (AC1 to AC10) over 100 hours, starting with initial temperatures ranging from 25 to 30°C. The reduction follows a steady linear decrease at 0.05°C per hour, representing a cooling effect or energy-saving operation. The initial temperatures for the 10 AC units were randomly distributed between 25 and 30°C. For example, AC1 started at 28.5°C, AC2 at 27.8°C, and AC10 at 25.2°C (these values are illustrative). Each AC unit exhibits a linear decrease in temperature over time, with the rate determined by the reduction factor. After 50 hours, AC1 cooled to approximately 26.0°C, while AC10 cooled to about 22.7°C.

By the end of 100 hours, the units' temperatures had reduced to their lowest points, ranging from approximately 20 to 25°C, depending on their starting values. The plot vividly illustrates the cooling trend for each unit. The consistent slopes highlight the uniform reduction rate, while the individual starting points reflect the variability in initial conditions. This simulation

https://doi.org/10.31436/iiumej.v26i3.3629

provides valuable insights into system behavior under energy-saving strategies, ensuring temperature targets are met over time.

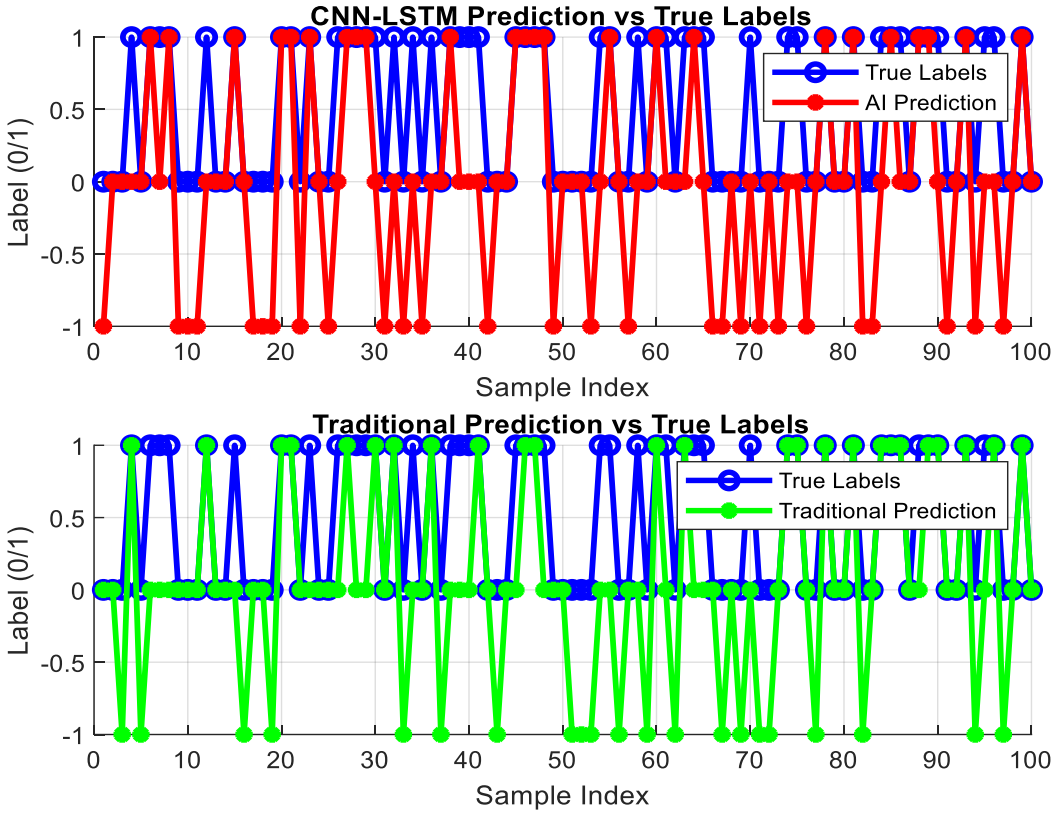


Figure 13. Accuracy and Precision Evaluation.

4.2.2. Reduced Energy Consumption in Data Centers

The plot in Figure 15 shows the reduction in energy consumption for data centers over 100 hours, starting from an initial consumption of 100 arbitrary units. The energy consumption decreased exponentially, with a reduction factor of about 2% per hour, as optimized through AI predictions. After 50 hours, energy consumption dropped to approximately 36.8 units, and by the end of 100 hours, it reached about 13.5 units. The curve demonstrates a steady and efficient decline in energy use, highlighting the impact of optimization techniques. This result emphasizes the significant energy savings achievable in data center operations while maintaining a continuous reduction trend over time.

4.2.3. Reduced Cooling Costs in Data Centers

The plot in Figure 16 illustrates the reduction in cooling costs, which began at 500 arbitrary units. The costs also followed an exponential decline, with a higher % reduction factor of 3% per hour. At the halfway mark (50 hours), the cooling costs decreased to around 111 units; by 100 hours, they reached approximately 25 units. The steeper slope compared to energy consumption reflects the faster rate of cost reduction, underscoring the economic benefits of AI-optimized cooling systems. These results reveal how enhanced cost-saving measures complement energy reduction efforts, leading to a more sustainable and cost-effective data center operation.

https://doi.org/10.31436/iiumej.v26i3.3629

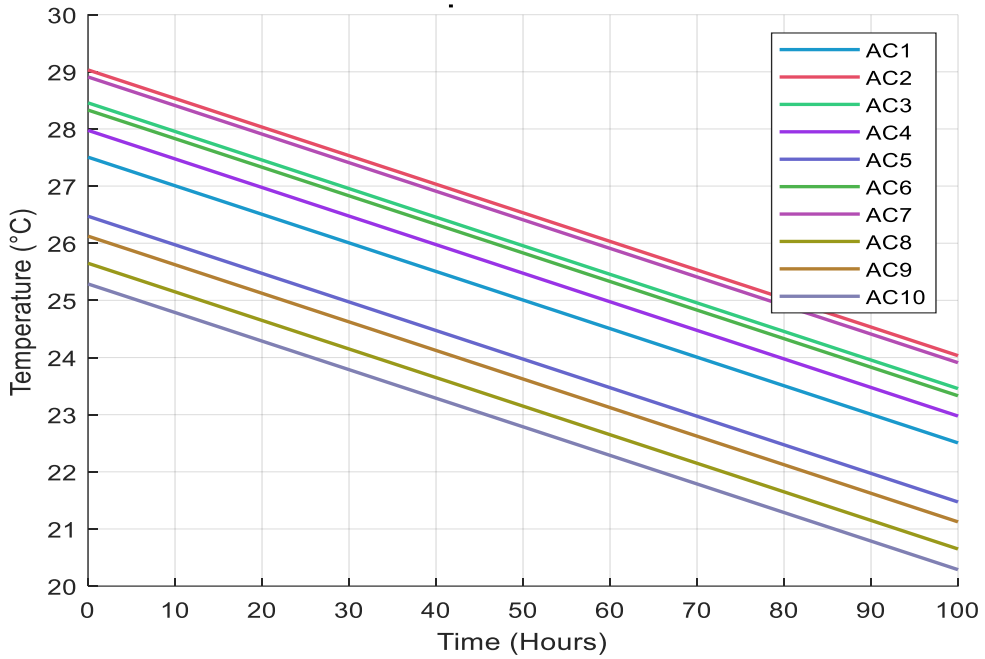


Figure 14. Reduce Temperature of AC1 to AC10

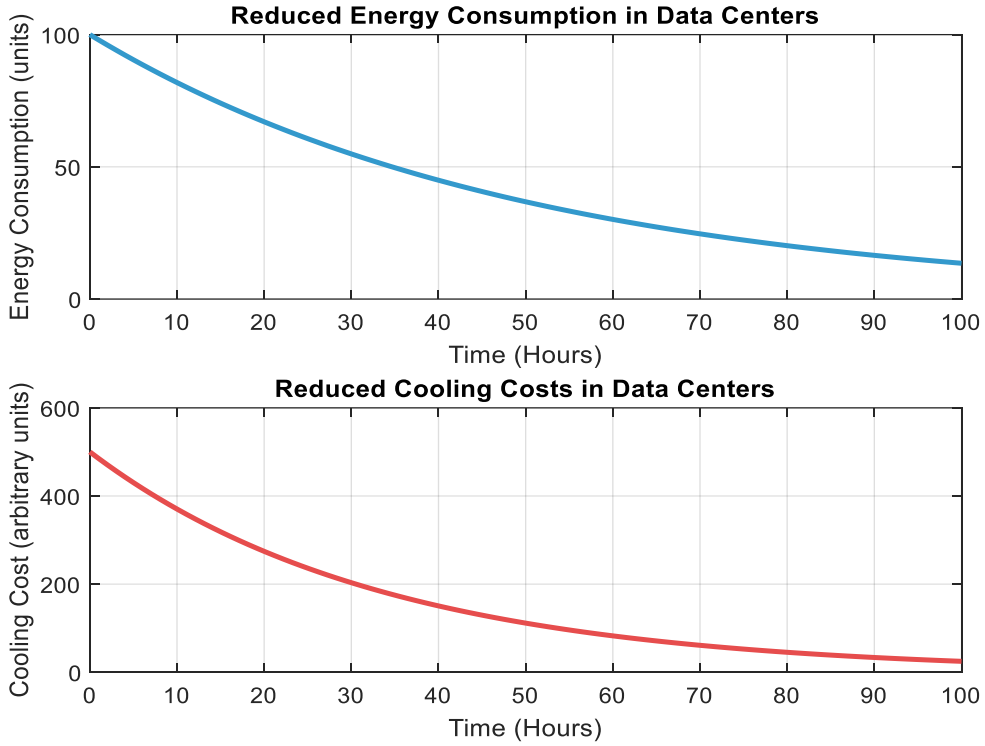


Figure 15. Reduced Energy Consumption.

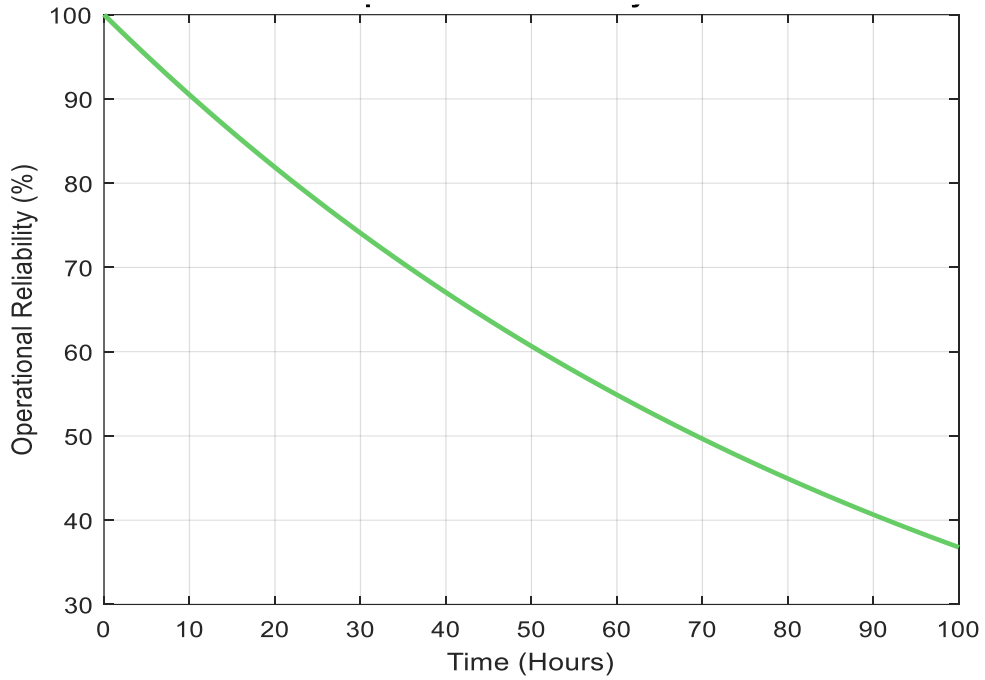


Figure 16. Enhanced Operational Reliability.

5. CONCLUSION

This study successfully developed and evaluated a hybrid CNN-LSTM predictive cooling system for real-time temperature management in data centers. The predictive capability ensures thermal stability, which underscores the system's ability to address the limitations of traditional temperature management methods by incorporating advanced machine learning techniques. Considering key performance metrics, the system reliability underlines the resilience of transient disruptions; the cooling efficiency demonstrates dynamic adaptability, while the model's quick response time highlights its capability to react promptly to heat changes at the center. Energy consumption and efficiency results reveal significant benefits of predictive cooling, while energy use aligned with operational demands and efficiency improvements of up to 7% during peak conditions. Achievement of high prediction accuracy, ranging between 85 and 95% validates the effectiveness of the hybrid model.

This research demonstrates the practical benefits of integrating machine learning into cooling system design, offering significant improvements in energy efficiency, operational stability, and environmental sustainability. The insights gained here pave the way for future advancements in predictive cooling technologies, emphasizing the critical role of data-driven solutions in optimizing performance and reducing energy costs in modern data centers.

REFERENCES

[1] Huang M, Liu Z, Tao Y. (2022) Mechanical fault diagnosis and prediction in IoT based on multi-source sensing data fusion. Simulation Modelling Practice and Theory, 2022 102:101981.

[2] Harris A, Yellen M (2024) Decision-making with machine prediction: Evidence from predictive maintenance in trucking. Retrieved from: https://adamharris380.github.io/files/HarrisYellen-JMP.pdf

[3] Cakir M, Akbulut A. (2022) A Bayesian deep neural network approach to seven-point thermal sensation perception. IEEE Access, 2022 10:5193-5206.

[4] Xia L, Liang Y, Leng J, Zheng P. (2023) Maintenance planning recommendation of complex industrial equipment based on knowledge graph and graph neural network. Reliability Engineering & System Safety, 2023 232:109068.

[5] Olabisi P, Ayegbokiki J. (2024) Review of Computational Optimization Techniques for the Coordination of Over-Current Relay Protection in Sub-Transmission Stations. Applications of Modelling and Simulation, ARQII Publication, http://arqiipubl.com/ams, eISSN 2600-8084. 2024, 8:162-173.

[6] Gao Y, Fang C, Zhang J. (2022) A spatial analysis of smart meter adoptions: Empirical evidence from the US data. Sustainability, 2022 14:1126.

[7] Ouadah A, Zemmouchi-Ghomari L, Salhi N. (2022) Selecting an appropriate supervised machine learning algorithm for predictive maintenance. The International Journal of Advanced Manufacturing Technology, 2022 119(11):4277–4301.

[8] Chicco D, Warrens MJ, Jurman G. (2021) The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE, and RMSE in regression analysis evaluation. Peer J. Computer Science, 2021 7, Article e623.

[9] Ahmed SF, Alam S, Hassan M, Rozbu MR, Ishtiak T, Rafa N, Mofijur M, Ali ABMS, Gandomi AH. (2023) Deep learning modelling techniques: current progress, applications, advantages, and challenges. Artificial Intelligence Review 2023 56:13521–13617 https://doi.org/10.1007/s10462-023-10466-8

[10] Sarker IH (2021) Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. SN Computer Science, 2021 2:420. https://doi.org/10.1007/s42979-021-00815-1

[11] Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang G, Cai J, Chen T. (2018) Recent advances in convolutional neural networks. Pattern Recognition (Elsivier), May 2018 77:354-377.

[12] Ilager S, Ramamohanarao K, Buyya R. (2019) ETAS: Energy and thermal-aware dynamic virtual machine consolidation in cloud data center with proactive hotspot mitigation. Concurrency and Computation: Practice and Experience, 2019, Article e5221. https://doi.org/10.1002/cpe.5221

[13] Ma H, Ding A. (2020) Method for evaluation on energy consumption of cloud computing data center based on deep reinforcement learning. Electric Power Systems Research, 2020 208, Article 107899.

[14] Ramasamy V, Thalavai Pillai S. (2020) An effective HPSO-MGA optimization algorithm for dynamic resource allocation in cloud environment. Cluster Computing, 2020 23(3):1711-1724.

[15] Stoffel P, Henkel P, Ratz M, Kumpel A, Muller D. (2023). Safe operation of online learning data driven model predictive control of building energy systems. Energy and AI. 13, 100296.

[16] Guo Q, He Z, Wang Z. (2024) Monthly climate prediction using deep convolutional neural network and long short-term memory. Scientific Reports, 2024 14, Article 17748.

[17] Hussain W, Hussain FK, Saberi M, Hussain OK, Chang E. (2018) Comparing time series with machine learning-based prediction approaches for violation management in cloud SLAs. Future Generation Computer Systems, 2018 89:464-477.

[18] Oh J, Kim Y. (2022) Job placement using reinforcement learning in GPU virtualization environment. Cluster Computing, 2022 23(3):2219-2234.

[19] Li J, Deng Y, Zhou Y, Zhang Z, Min G, Qin X. (2022) Towards thermal-aware workload distribution in cloud data centers based on failure models. IEEE Transactions on Computers, 2022 1.

[20] Rafi SH, Nahir-Al-Masood, Deeba SR, Hossain E. (2021) A short term load forecasting method using integrated CNN and LSTM network. IEEE ACCESS, 9 32436-32448. https://doi.org/10.1109/ACCESS.2021.3060654

[21] Elmaz F, Eyckerman R, Casteels W, Latré S, Hellinckx P. (2021) CNN-LSTM architecture for predictive indoor temperature modeling. Building and Environment, 2021 206:108327.

[22] Bekkar A, Hssina B, Douzi S, Douzi K. (2021) Air-pollution prediction in smart city, deep learning approach. Journal of Big Data, 8, Article 161. https://doi.org/10.1186/s40537-021-00563-5

[23] Cao J, Harrold D, Fan Z, Morstyn T, Healey D, Li K. (2020) Deep reinforcement learning-based energy storage arbitrage with accurate lithium-ion battery degradation model. IEEE Transactions on Smart Grid, 2020 11:4513–4521. doi: 10.1109/tsg.2020.2986333

[24] Balakumar P, Vinopraba T,.Chandrasekaran K. (2023) Deep learning based real-time demand side management controller for smart building integrated with renewable energy and energy storage system, Journal of Energy Storage, 2023 58:106412. doi: 10.1016/j.est.2022.106412.

https://doi.org/10.31436/iiumej.v26i3.3633

MODELLING OF FLEX SENSOR RESISTANCE VALUE FOR FINGERSPELLING RECOGNITION PROTOTYPE USING SENSOR-BASED SIGNAL

MUHAMMAD HAZMIRUL HAQIM HAZRI, ZARINA MOHD NOH*, NORHIDAYAH MOHAMAD YATIM, NIK MOHD ZARIFIE HASHIM

Fakulti Teknologi dan Kejuruteraan Elektronik dan Komputer (FTKEK), Universiti Teknikal Malaysia Melaka (UTeM), Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia

**Corresponding author: zarina.noh@utem.edu.my*

(Received: 3 March 2025; Accepted: 8 May 2025; Published online: 9 September 2025)

ABSTRACT: Individuals with hearing and speech impairments often rely on sign language to communicate with each other. While the hearing and speech impairment community can understand sign language to communicate with each other, individuals with normal hearing generally do not understand sign language. However, communication between the hearing and speech-impaired individual and a normally hearing individual can be made possible by fingerspelling with the help of a fingerspelling recognition system. Current research in fingerspelling recognition systems aims to improve the system's accuracy, rather than focusing on the ability of the system to convert the fingerspelling for real-time communication. The developed system can be computationally complex when the research aims to improve accuracy, resulting in unsuitability for real-time communication due to the extensive processing needed and oversized non-wearable hardware. Hence, this paper presents one of the options in developing a fingerspelling recognition system prototype that aims for ease in real-time communication using sensor-based signals. The prototype comprises a hand glove equipped with five flex sensors, designed to recognize and convert the hand gestures into the fingerspelled letters. Each flex sensor is fitted along the length of each finger to capture the gestured fingerspelling, measuring the changes in the flex sensor's resistance as it bends. The flex sensor resistance value was then modelled for fingerspelling recognition through demonstration on five users in the preliminary characterization, and later by ten users to obtain the system accuracy. An Arduino MEGA 2560 platform processed the sensor's data, which then transmits the recognized fingerspelled letters to an OLED display. The average recognition accuracy recorded for the prototype is 76.15% which is 15.38% higher than the preliminary characterization, indicating the possibility of using sensor-based signals in a fingerspelling recognition system for real-time communication. This system allows individuals with hearing and speech impairment to make various hand gestures while wearing the hand glove, which will be recognized as the corresponding fingerspelled letters for easy communication with individuals with normal hearing.

ABSTRAK: Individu dengan kekurangan pada pendengaran dan pertuturan biasanya bergantung kepada bahasa insyarat bagi berkomunikasi antara satu sama lain. Walaupun komuniti dengan kekurangan pada pendengaran dan pertuturan boleh memahami bahasa isyarat bagi berkomunikasi antara satu sama lain, individu dengan pendengaran normal secara amnya tidak memahami bahasa isyarat. Walau bagaimanapun, komunikasi antara individu kekurangan pada pendengaran dan pertuturan, dengan individu normal boleh diwujudkan melalui ejaan jari dengan bantuan sistem pengecaman ejaan jari. Kajian terkini dalam sistem pengecaman ejaan jari bertujuan bagi meningkatkan ketepatan sistem, bukan memfokuskan kepada kebolehan sistem dalam menterjemah ejaan secara masa nyata. Apabila tujuan kajian

https://doi.org/10.31436/iiumej.v26i3.3633

tersebut adalah bagi meningkatkan ketepatan, sistem yang terbina boleh menjadi sangat kompleks dalam pengiraannya, mengakibatkan sebuah sistem yang tidak sesuai bagi tujuan komunikasi masa nyata kerana memerlukan proses berlebihan dan perkakasan bersaiz besar yang tidak boleh dipakai. Oleh itu, kajian ini memberi pilihan dalam membangunkan sebuah prototaip sistem pengecaman ejaan jari, bertujuan memudahkan komunikasi masa nyata menggunakan isyarat berasaskan penderia. Prototaip ini terdiri daripada sebuah sarung tangan yang dilengkapi dengan lima penderia lentur, direka bagi mengenalpasti huruf yang dieja jari pada gaya tangan. Setiap penderia lentur dimuatkan sepanjang setiap jari bagi mengesan ejaan jari yang digayakan. Ukuran dibuat melalui perubahan rintangan penderia lentur apabila ianya dibengkokkan. Nilai rintangan penderia lentur ini kemudiannya dimodelkan bagi pengecaman ejaan jari melalui demonstrasi ke atas lima pengguna dalam perincian awal, dan seterusnya oleh sepuluh pengguna bagi mendapatkan ketepatan sistem. Data penderia tersebut diproses menggunakan platform Arduino MEGA 2560, dan huruf ejaan jari yang telah dikenal pasti dipaparkan ke paparan OLED. Purata ketepatan pengecaman yang direkodkan bagi prototaip ini adalah 76.15% di mana 15.38% lebih tinggi dari perincian awal, menunjukkan potensi penggunaan isyarat berasaskan penderia dalam sistem pengecaman ejaan jari bagi komunikasi masa nyata. Sistem ini membolehkan individu kekurangan pada pendengaran dan pertuturan membuat pelbagai gaya tangan sewaktu memakai sarung tangan ini, kemudian huruf ejaan jari yang sepadan akan dikenalpasti bagi memudahkan komunikasi bersama individu normal.

KEYWORDS: *Fingerspelling, Flex Sensor, Hand Gesture, Sign Language.*

1. INTRODUCTION

A forecast by the World Health Organization (WHO) stated that one in four people worldwide will have hearing impairment by 2050 [1]. Focusing on Malaysia, the Department of Social Welfare reported that about 46,127 individuals are currently affected by hearing impairment [2]. In general, interactions between individuals with hearing impairment and individuals with normal hearing have always been a troublesome task. This is because not everyone can comprehend sign language. Around the world, the use of sign language varies according to countries such as 'Bahasa Isyarat Malaysia' (BIM) in Malaysia [3], American Sign Language (ASL) in United States [4], Mexican Sign Language (LSM) in Mexico [4], and 'Bahasa Isyarat Indonesia' (BISINDO) in Indonesia [5]. Each sign language has its own uniqueness in either using both or one hand, but there are also some similarities in these sign language gestures, depending on their origin. Therefore, researchers in the area have been developing sign language recognition systems that are customized to the sign language in use, such as FingerSpeller [6], GestureSpeak [7], SpellRing [8], and SHWASI [9].

FingerSpeller recognized ASL fingerspelling using a wearable keyboard with several sensors embedded on rings [6]. The user must wear the ring on each finger, where hand gestures will be translated to text. While the average accuracy of the system can achieve up to 91%, each user had to personally collect their fingerspelled inputs to train the system for the customized recognition system. This requires additional procedures from the user to recheck their collected data to ensure accurate data was recorded as the training input. SpellRing also recognized ASL fingerspelling, but it utilized active acoustic sensing to detect hand gestures by a ring worn on the thumb, and a gyroscope [8]. The system can achieve up to 89.89% accuracy, but it needs a larger dataset in the recognition system's pre-trained model to be user-independent. Furthermore, the system needs to be physically connected to a personal computer for recognition processing, which can be bulky and impractical for real-time communication. On the other hand, SHWASI requires users to make specific hand gestures according to a pre-defined set of movements using a hand glove attached with flex sensors [9]. The system provides three modes for the recognition options to translate into phrases, letters, or alphabets

other than letters. While the system's accuracy can be up to 79%, users had to learn a new set of pre-defined finger gestures to use the system. It can be challenging for users to learn pre-defined finger gestures just to use the system, as they are used to their daily standard sign language. GestureSpeak is an ASL fingerspelling recognition system developed to achieve a high recognition accuracy through a computationally efficient algorithm for faster response time [7]. The system requires data training using images of various hand gestures and can achieve up to 98.42% accuracy. While the accuracy is the highest compared to the other sensor-based systems mentioned, GestureSpeak is a vision-based system not yet deployed in hardware for real-time communication, as the reported result was based on an experimental setup [7]. Still, the system developed by researchers in the area proved that researchers are now moving towards improving the accuracy of the fingerspelling recognition system while aiming for the system to be utilized in real-time communication.

Hence, this paper explores one of the possibilities: developing a fingerspelling recognition system with acceptable accuracy that is enough to be used in real-time communication. This work aims to bridge the communication gap between hearing impaired individuals and normal hearing individuals using sensor-based signals in the fingerspelling recognition system. The proposed work in this paper incorporates flex sensors into a hand glove to capture finger movements as a sign language interpreter. It provides real-time output that displays the recognized letter. It is a promising initiative that can significantly enhance accessibility and communication for the hearing and speech-impaired community. The following section in this paper shares the background review in the area, while the following section explains the modelling and development of the fingerspelling recognition prototype. The following section presents the prototype demonstration and analysis of the system's accuracy in recognizing the letters. The last section concludes with the findings of this paper and future work that can be done in the area.

2. BACKGROUND STUDY

Sign language recognition systems can bridge the communication gap between sign language users and non-users by identifying and converting sign language into frequently spoken languages, allowing for seamless interaction and understanding [10]. They also play an important role in empowering the hearing-impaired community by giving them the ability to communicate more freely and effectively with a wider community. In addition, hearing impaired individuals can communicate through sign language more easily using fingerspelling.

2.1. Fingerspelling Recognition System

Fingerspelling involves representing the alphabet with one hand. It helps introduce new phrases and concepts, particularly when discussing words, names, or places without agreed-upon signs within the sign language community [11]. In fact, all words used by normal hearing individuals can be derived through fingerspelling. As such, the normal hearing community can also understand the messages conveyed by the hearing and speech impairment community without knowing the sign language beforehand, if there is a system that can translate the fingerspelling into text display. Such a system is also called a fingerspelling recognition system. In Malaysia, the fingerspelling practices by the hearing and speech impairment community follow the BIM as shown in Figure 1 [12]. The fingerspelling in BIM is gestured by using one hand for all letters, with exceptions for letters ‘J’ and ‘Z’ that are accompanied by some specific hand movement.

https://doi.org/10.31436/iiumej.v26i3.3633

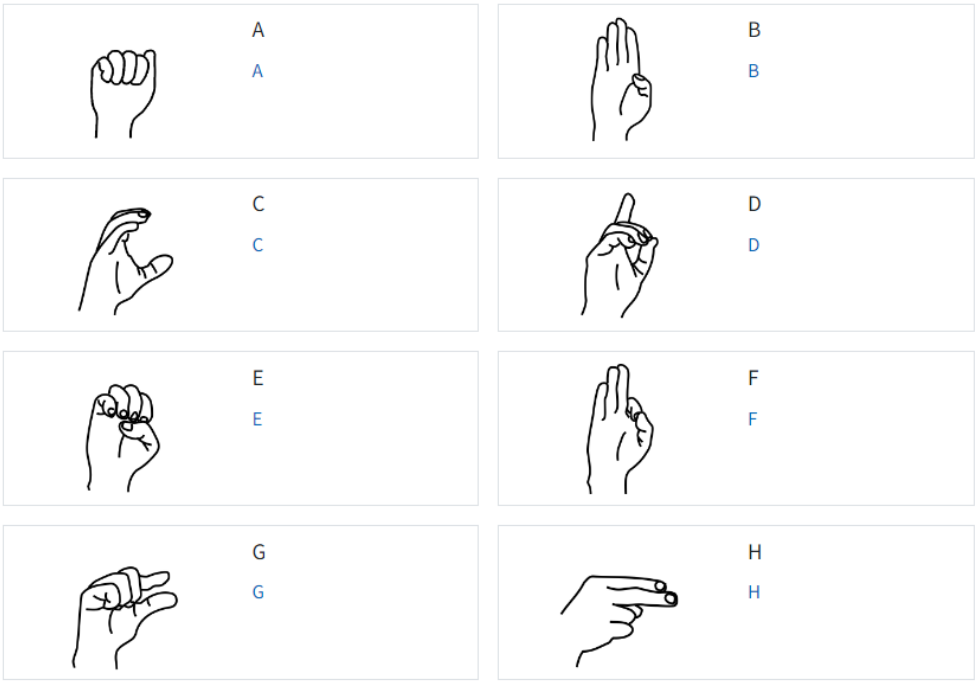


Figure 1. Sign language gestures for some of the letters in 'Bahasa Isyarat Malaysia' (BIM) [12].

A fingerspelling recognition system can either be developed by a vision-based or sensor-based system [13]. While a vision-based system uses images to convert the fingerspelling gesture into text, a sensor-based system uses readings obtained from sensors to convert the fingerspelling gesture [14]. As such, vision-based systems usually require a high-performance computing platform for image processing that uses high-end hardware and software in their developmental phase. This is because a vision-based system is sensitive to factors affecting the image properties, such as surrounding illumination, image resolution, and colour variation, to achieve a high-accuracy system [5]. This resulted in an impractical system more suitable for the research and developmental phase than for real-time communication. In domestic use for real-time communication, a sensor-based system is an appropriate alternative as it can utilize the off-the-shelf (OTS) components that are wearable. Although it is wearable, a sensor-based system can help translate fingerspelled gestures with accuracy that is enough for real-time fingerspelling recognition [6][8][9]. The following section will further explain insights into the sensor-based approach for the fingerspelling recognition system.

2.2. Sensor-based Fingerspelling Recognition System

In a broader sense, sensor-based sign language recognition systems use sensors to capture and interpret the gestures and movements made during sign language communication [15]. Such a system relies on sensor data, such as flex sensors, accelerometers, or wearable devices, to detect and analyze the hand and arm movements involved in sign language gestures. By processing the sensor data, sensor-based systems can recognize and translate signs and gestures into text or speech, facilitating communication for individuals with hearing and speaking impairments. An example of a sensor-based sign language interpreter system is shown in Figure 2.

The system in Figure 2 uses flex sensors attached to the fingers to detect the bending of the fingers when the user is gesturing the sign language. The inertial measurement unit (MPU 6050) is also used to measure the dynamic and static gestures of the hand from the gyroscope

https://doi.org/10.31436/iiumej.v26i3.3633

and accelerometer readings. The data obtained from the flex sensors and the inertial measurement unit are processed by the Raspberry Pi as the processor platform [16]. Once the sign language gestures had been translated and identified, the Raspberry Pi sends the data for display through the LCD and is voiced out by the speaker.

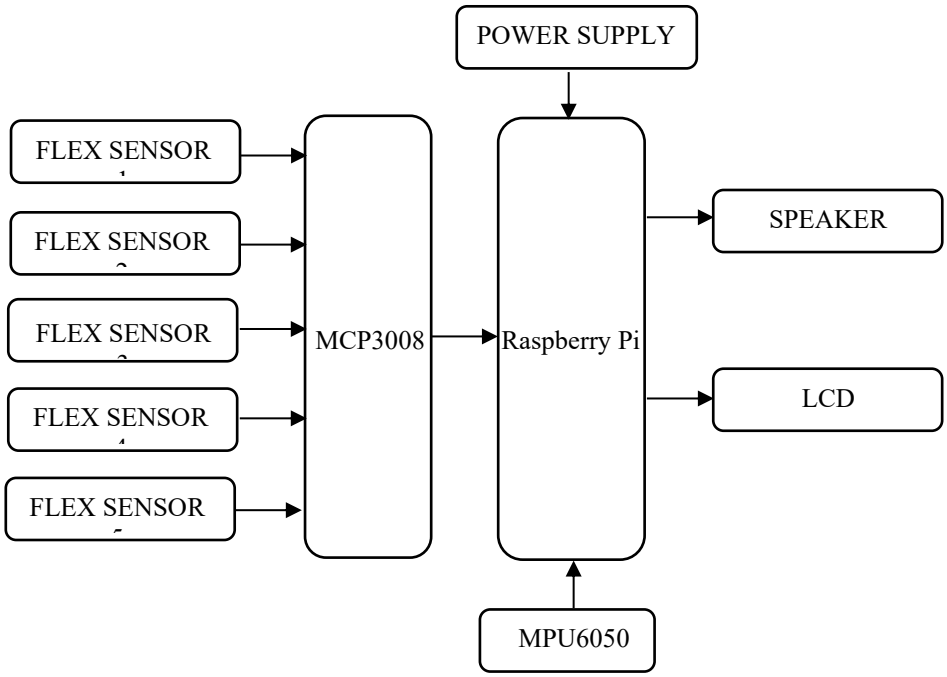


Figure 2. Sensor-based sign language interpreter system [16].

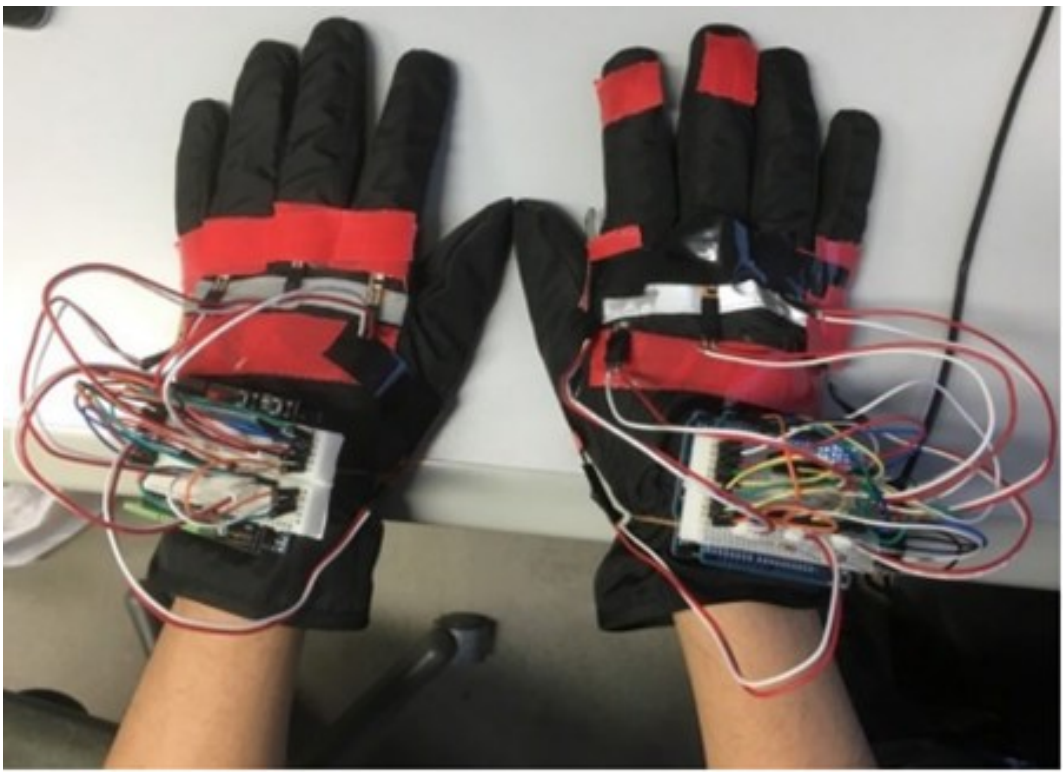


Figure 3. Prototype of sensor-based sign language interpreter glove system using accelerometer [17].

https://doi.org/10.31436/iiumej.v26i3.3633

Two commonly used sensors in sensor-based sign language recognition systems are flex sensors and accelerometers. A flex sensor is used to detect finger bending (usually attached to a glove) as the resistance of the sensor changes when the finger is bent, allowing the system to capture the changes in movements involved in sign language [15]. On the other hand, the accelerometer helps detect dynamic gestures like the alphabet's J' and 'Z' in BIM and ASL. These alphabets involve hand movements that are not easily captured by static sensors like flex sensors. The accelerometer, which measures the acceleration of the hand or fingers, helps to detect the dynamic movements involved in these two alphabets. Figure 3 shows an example of a sign language recognition system prototype using the accelerometer [17]. The accelerometer in Figure 3 was used to detect the hand motion in translating sign language.

3. MODELLING AND DEVELOPMENT OF FINGERSPELLING RECOGNITION PROTOTYPE

Based on the background study in the previous section, the prototype of the fingerspelling recognition system in this paper was developed as shown in Figure 4, the block diagram. Five flex sensors were connected to the Arduino MEGA 2560 platform as the input that sensed each finger's bending signal. These flex sensors were fixed on the hand glove that the user needs to wear. An accelerometer was added as a sensor to detect the movement of the letters 'J' and 'Z'. The output of the prototype includes green and red LEDs, a buzzer, and an OLED display. The green and red LEDs served as indicators for valid or invalid sensor readings, while the buzzer would be triggered once a letter had been recognized from the valid sensor reading. The OLED display would show the letters detected from the sensor readings and the analog-to-digital (ADC) readings detected from the flex sensors.

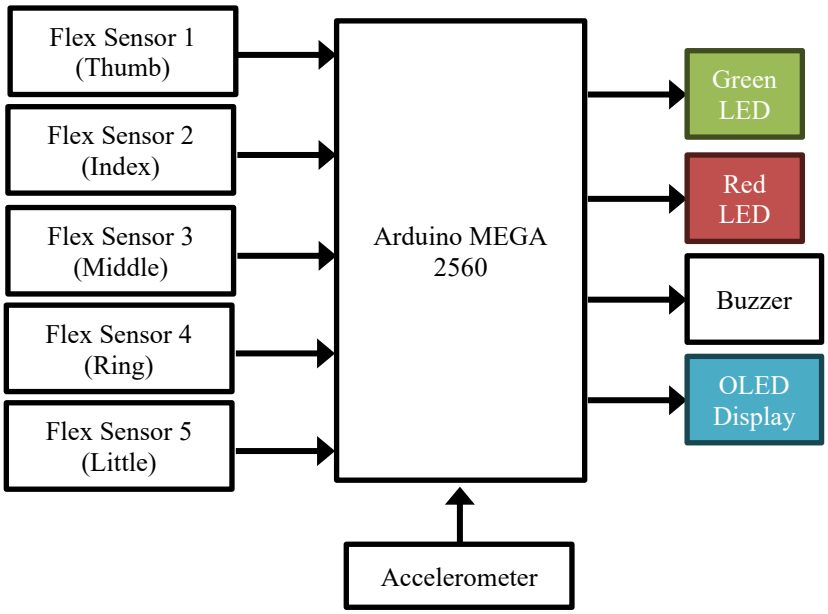


Figure 4. Block diagram of the developed fingerspelling recognition prototype.

The schematic diagram of the prototype is shown in Figure 5. The two LEDs and buzzer are connected to the digital pins (D8, D9 and D13) of the Arduino MEGA 2560, while the five flex sensors can be seen connected to the analog pins of the board (A0, A1, A2, A3 and A4). Each component was accompanied by suitable resistors to limit the current flow through the pins. The OLED display and accelerometer both use the I2C pins on the Arduino MEGA 2560, where the access time between the two components was controlled through the program code

https://doi.org/10.31436/iiumej.v26i3.3633

to enable valid communication between the two components. The printed circuit board (PCB) configured for the prototype is divided into two boards, where one PCB is for the Arduino MEGA 2560 and its peripherals, other than the accelerometer, while the other PCB is for an isolated accelerometer.

The PCB developed for the Arduino MEGA 2560 connections is shown in Figure 6. The top view in Figure 6 shows the green and red LEDs, buzzer, OLED display, and headers for flex sensor connections. There are also headers for future system expansion on the PCB, as labeled in Figure 6. The bottom view in Figure 6 shows the circuit route between the inputs and outputs to the Arduino MEGA 2560 pins. The input and output components of the prototype were connected to the PCB board via male or female connectors, respectively. The PCB's corresponding top and bottom views were also labeled in Figure 6 to show the relationship between both views concerning their headers and pins.

The other PCB that housed the accelerometer can be seen in Figure 7. The accelerometer was connected to an isolated PCB because it needs to be attached to the hand glove for hand movement detection. As the accelerometer shared the I2C pins with the OLED display in its connection to the Arduino MEGA 2560, the programming code controls access to the I2C serial communication. The isolated PCB for the accelerometer in Figure 7 can be seen in Figure 8, attached to the hand glove. This is because hand movement can be precisely sensed if the accelerometer is placed on top of the user's hand, rather than housed on the same PCB as the Arduino MEGA 2560 platform.

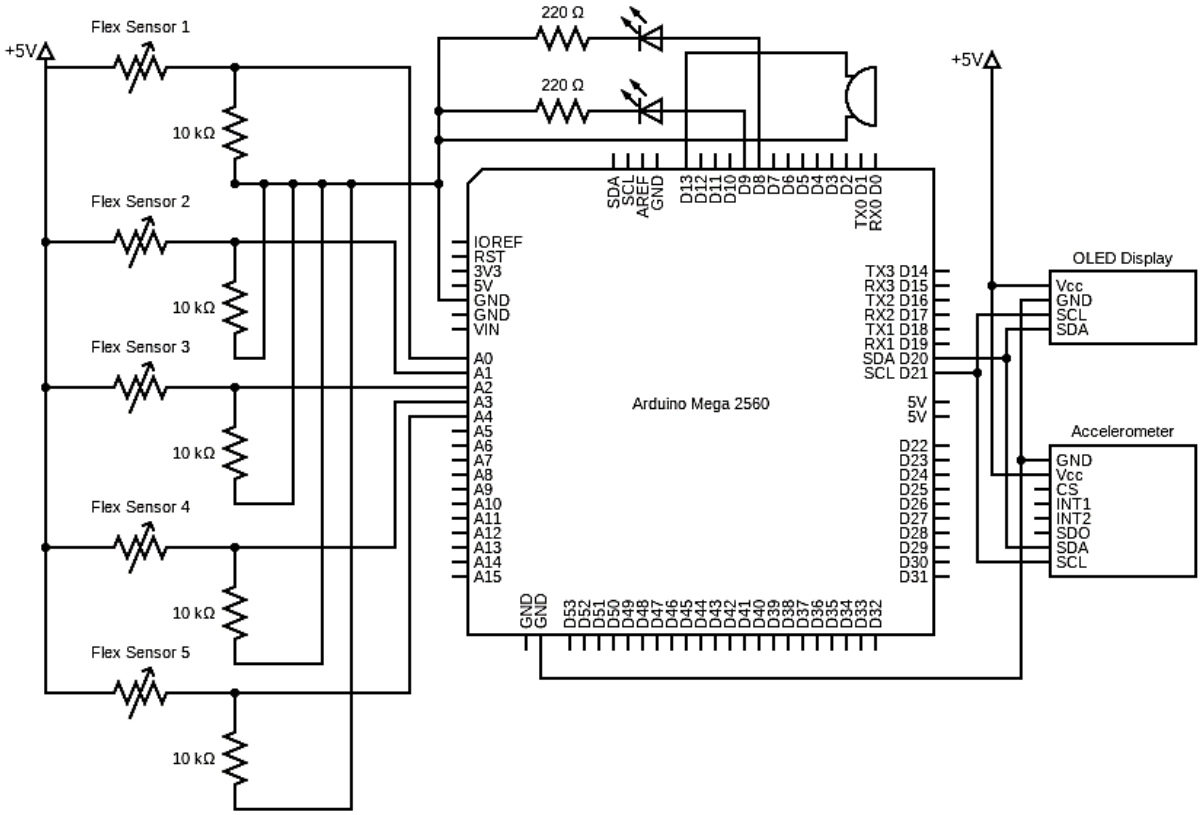
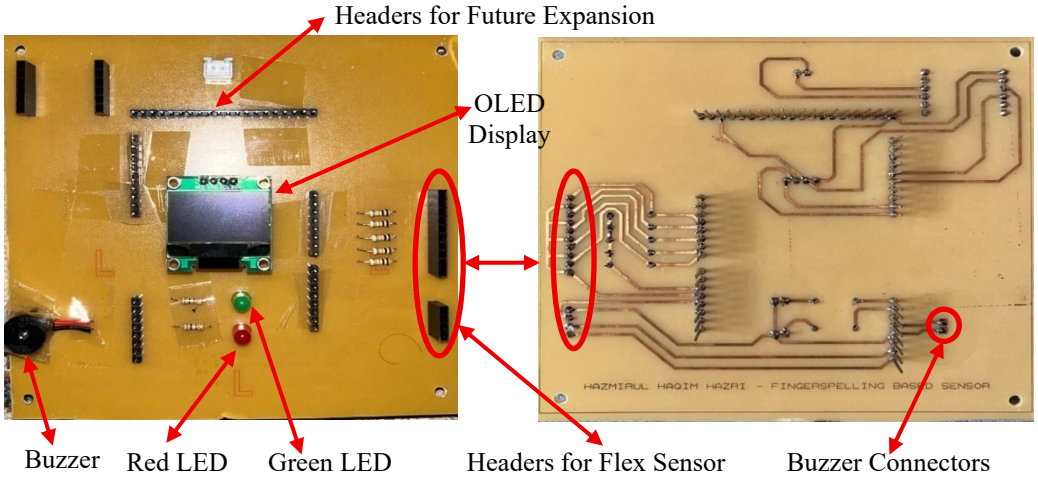


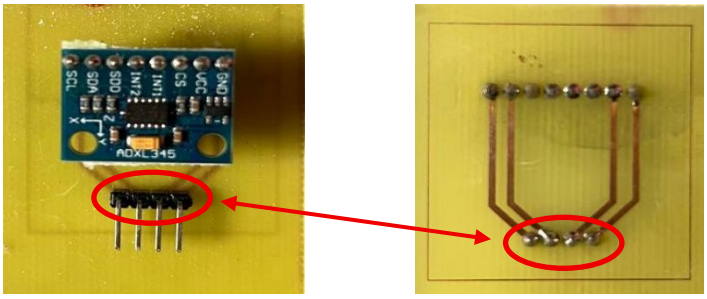
Figure 5. Schematic diagram of the fingerspelling recognition prototype.

https://doi.org/10.31436/iiumej.v26i3.3633



(a) Top view (b) Bottom view

Figure 6. Top and bottom views of the prototype's PCB, with labels showing the corresponding pins on both views.



(a) Top view (b) Bottom view

Figure 7. Top and bottom views of the accelerometer's isolated PCB, with labelling to show the top and bottom view connections.

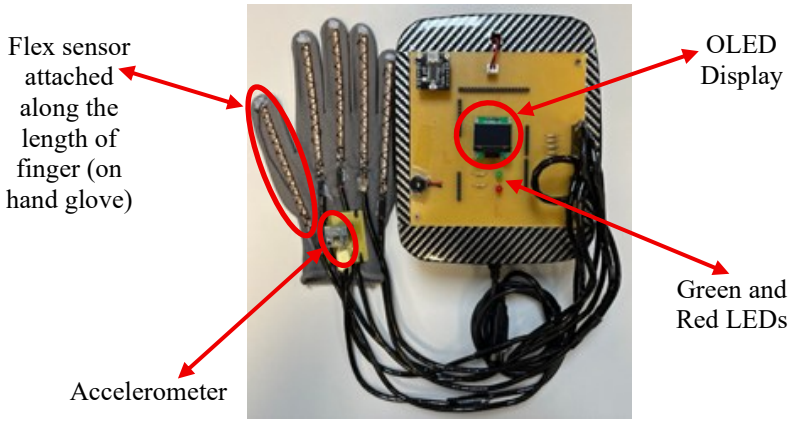


Figure 8. Final prototype of the developed fingerspelling recognition system.

Figure 8 shows the final prototype of the developed fingerspelling recognition that combines both PCBs, as shown in Figs. 6 and 7. The hand glove in Figure 8 was attached with the flex sensors along the length of each finger, together with the isolated PCB that housed the accelerometer. Once a hearing impairment user wears the hand glove and gestures a

https://doi.org/10.31436/iiumej.v26i3.3633

fingerspelling, the flex sensors will sense the resistance value produced by the finger's bending, where the readings are converted to their ADC values by the Arduino MEGA 2560. Through the fingerspelling recognition processing in the Arduino MEGA 2560 program code, the OLED display in Figure 8 will display the recognized letter for a normal hearing individual to understand the fingerspelling. In Figure 8, the Arduino MEGA 2560 was housed in the enclosed box. A power bank will power the prototype to ease its mobility and enable its portability. Figure 9 shows the steps involved in modeling the flex sensor's resistance value for the prototype, which consists of two main processes: preliminary characterization and improved characterization.

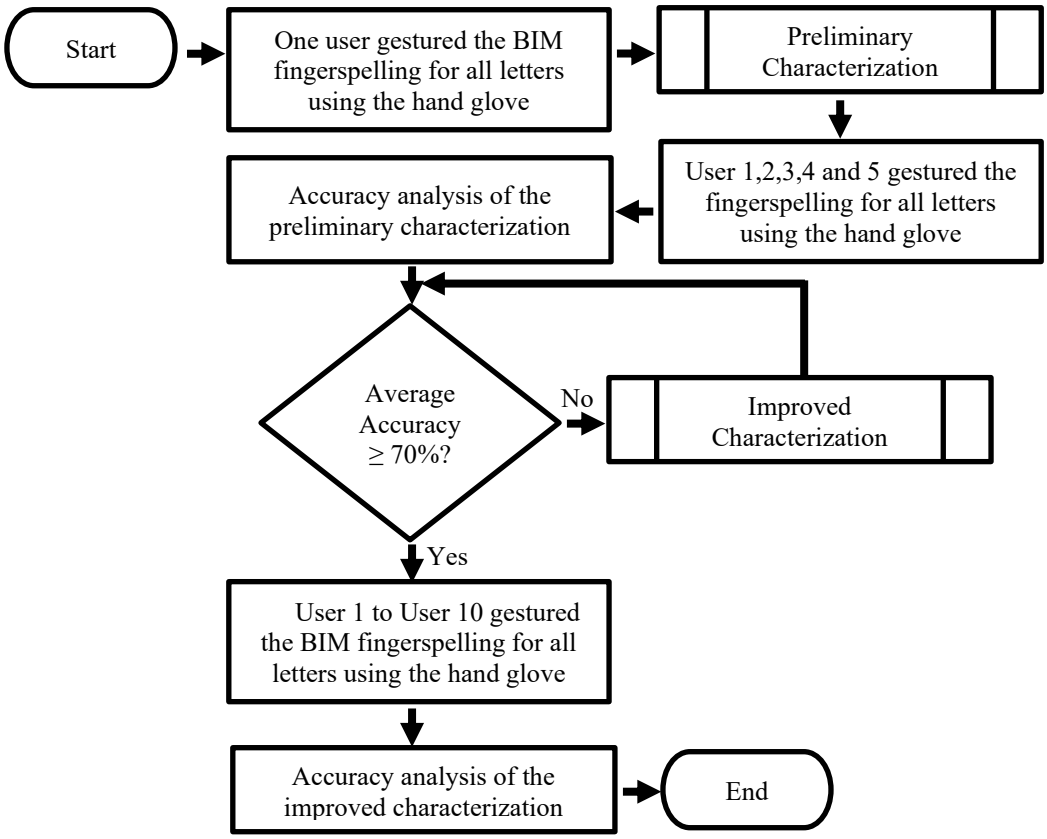


Figure 9. Steps involved in the prototype's modeling of flex sensors' resistance value.

In the preliminary characterization process shown in Figure 9, the flex sensors' resistance produced by one user when the BIM fingerspelled each letter was gestured is recorded in its converted ADC value. The converted ADC values from this user act as the threshold values for each BIM fingerspelled letter in the preliminary characterization. The threshold values were used in a conventional if-else algorithm for the fingerspelling recognition processing to map the hand gestures to the fingerspelled letters. The threshold values in the preliminary characterization for the fingerspelling recognition processing are shown in Table 1. Afterwards, the prototype was tested by five users, and their flex sensors' resistance values were recorded to determine the accuracy of the prototype in the preliminary characterization. In this phase, the threshold values in the conventional if-else algorithm of the fingerspelling recognition processing were based on the threshold values given in Table 1.

https://doi.org/10.31436/iiumej.v26i3.3633

Table 1. ADC threshold values for the preliminary characterization in the fingerspelling recognition processing

	ADC Threshold Value for Each Flex Sensor based on Finger				
Letters	**Thumb**	**Index**	**Middle**	**Ring**	**Little**
A	433	253	272	283	224
B	316	486	511	501	446
C	369	386	379	420	427
D	343	469	305	336	276
E	289	261	277	308	249
F	286	290	476	494	440
G	432	418	254	268	199
H	322	476	479	304	268
I	288	246	280	336	423
J	367	306	333	367	446
K	331	481	502	342	277
L	431	468	281	293	248
M	296	264	316	306	228
N	364	270	282	291	221
O	343	344	337	380	387
P	397	480	482	431	202
Q	456	443	299	311	247
R	355	477	501	341	295
S	331	279	330	370	268
T	399	274	305	313	267
U	356	481	512	354	309
V	350	483	509	383	333
W	336	491	507	501	298
X	334	444	289	307	261
Y	342	333	372	394	448
Z	376	488	321	363	296

Based on the accuracy analysis of the five users in the preliminary characterization, the threshold values of the fingerspelling recognition processing were enhanced in the improved characterization process if the average recognition accuracy was less than 70%. The average recognition accuracy value of 70% was chosen as a proof-of-concept for this prototype's potential as a fingerspelling recognition system. In the improved characterization process, the threshold values of the fingerspelling recognition processing were enhanced based on the recorded flex sensors' ADC values of the five users in the preliminary characterization for each letter. The improvement was made by focusing on the letters that achieved an average accuracy of less than 70%, where the threshold values were increased according to the highest recorded ADC values of the five users. The new threshold values (meaning the highest recorded ADC values concerning letters that achieved less than 70% average accuracy) will be utilized in the conventional if-else algorithm of the fingerspelling recognition processing in the improved characterization process.

At this time, the same five users will test the prototype repeatedly until the average accuracy of the system is 70% or higher. This is because, every time changes were made to the threshold values in the conventional if-else algorithm of the fingerspelling recognition processing, the average accuracy according to each letter recognition also changed. In turn, it is not guaranteed that once the threshold values of the focused letters are changed, the accuracy of the letters where the threshold values remain unchanged will stay the same. As a result, the prototype had to be tested repeatedly in the improved characterization process to get the best threshold values that can compromise the prototype's potential in recognizing the fingerspelled letters, such that its average accuracy achieved 70% at most.

https://doi.org/10.31436/iiumej.v26i3.3633

Once the prototype's average accuracy hit 70% or higher, the utilized threshold values were fixed in the conventional if-else algorithm of the fingerspelling recognition processing. The latest fixed threshold values in the improved characterization for the fingerspelling recognition processing are shown in Table 2. By utilizing the threshold values in Table 2, the prototype was then tested by ten users to get its final average accuracy. The final average accuracy will then be analyzed for each letter's recognition to conclude the developed prototype's potential as a fingerspelling recognition system that utilizes the sensor-based signal.

Table 2. ADC threshold values for the fingerspelling recognition processing following the improved characterization process, once the prototype accuracy is higher than 70%

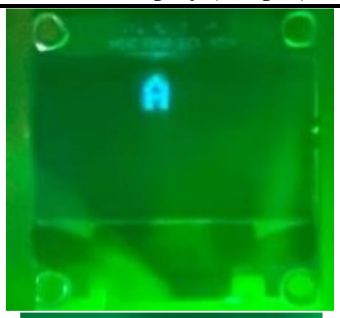
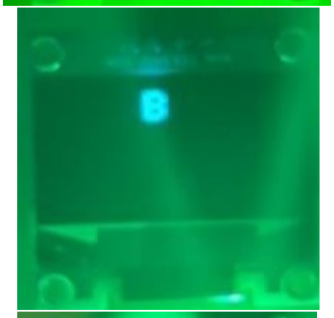
	ADC Threshold Value of Each Flex Sensor based on Finger				
Letters	**Thumb**	**Index**	**Middle**	**Ring**	**Little**
A	403	250	254	287	209
B	320	479	482	488	443
C	387	351	346	357	377
D	341	467	311	279	260
E	314	271	283	300	243
F	340	281	478	486	439
G	411	390	245	257	206
H	373	474	477	315	252
I	323	299	300	327	437
J	320	315	317	340	438
K	361	477	472	282	236
L	448	452	252	278	224
M	328	261	282	284	223
N	364	252	271	241	216
O	323	308	316	307	293
P	368	439	361	307	227
Q	427	439	254	290	213
R	328	468	466	234	242
S	290	275	274	277	209
T	374	288	249	270	208
U	319	425	479	282	289
V	318	422	485	324	285
W	330	418	483	479	270
X	318	300	256	258	212
Y	441	368	304	337	438
Z	322	384	271	275	216

4. RESULT AND ANALYSIS

Table 3 demonstrates the prototype's inputs and outputs when a user is gesturing five letters: 'A', 'B', 'C', 'D', and 'E'. The letters were gestured using the BIM fingerspelling, where the resistance value of the flex sensors on the hand glove was recorded as the prototype's input. These inputs, which were translated into their ADC values, are processed by the conventional if-else algorithm in the fingerspelling recognition processing to recognize the fingerspelled letters. The prototype then displays the recognized letter on the OLED display for real-time communication, as shown in Table 3.

https://doi.org/10.31436/iiumej.v26i3.3633

Table 3. Demonstration of the prototype operation when a user was using BIM fingerspelling for the letters 'A' to 'E'

Letter	Hand Gesture (Input)	OLED Display (Output)
A		
B		
C		
D		
E		

The average recognition accuracy of the prototype in the preliminary characterization for five users is given in Figure 10, where each user was labeled as *User 1*, *User 2*, *User 3*, *User 4,* and *User 5*. It can be seen in Figure 10 that the highest average accuracy is up to 88.46%,

https://doi.org/10.31436/iiumej.v26i3.3633

which was recorded for User 3, while the lowest average accuracy is 42.31% for *User 2*. The difference in average accuracy for each user is due to the variation in the length of each user's fingers. When the length of fingers differed for each user, the flex sensors would record different resistance values when bent, depending on the placement and fit of the hand glove on the user's hand. Hence, the choice of threshold values in the fingerspelling recognition processing plays a vital role in increasing or decreasing the average recognition accuracy of the prototype.

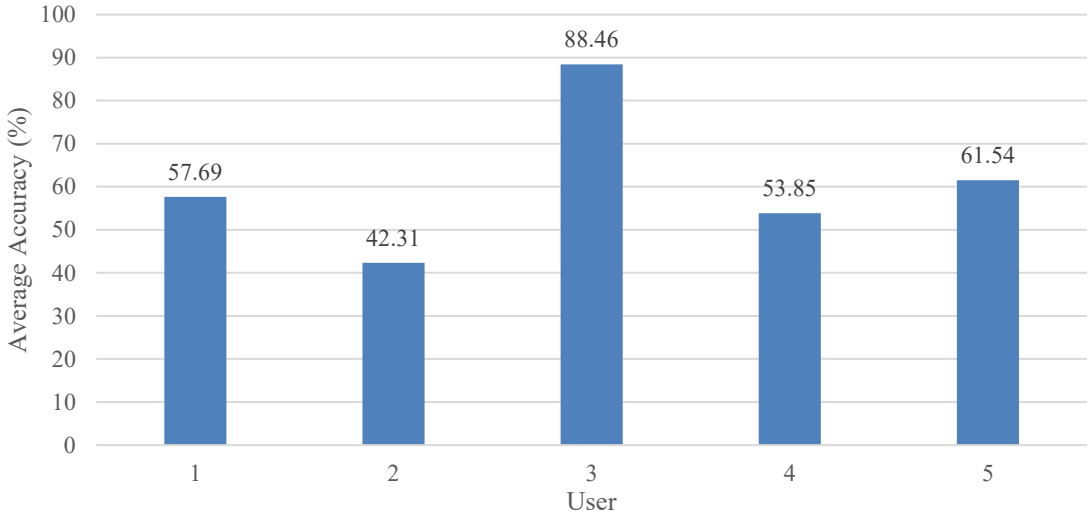


Figure 10. Average accuracy of the fingerspelling recognition system for five users using preliminary characterization values.

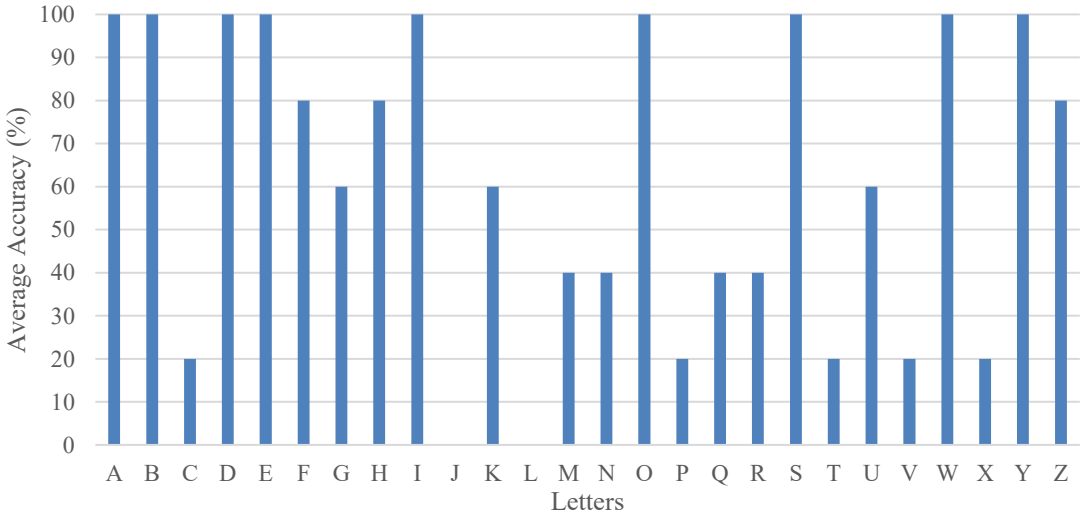


Figure 11. Average accuracy of the fingerspelling recognition system for each letter recognition using preliminary characterization values.

Figure 11 shows the average accuracy of the prototype in recognizing each letter for five users in the preliminary characterization. The average recognition accuracy of each letter shown in Figure 11 ranges from 0% to 100%. For letters such as 'A', 'B', 'D', 'E', 'I', 'O', 'S', 'W', and 'Y', the average recognition accuracy is 100%. The lowest average accuracy of 0% was recorded for letters 'J' and 'L', indicating that the prototype cannot recognize these two letters in the preliminary characterization. The average recognition accuracy for the rest of the letters varied between 20% to 80%. Based on the average accuracy in Figure 11, the total average

https://doi.org/10.31436/iiumej.v26i3.3633

accuracy of the fingerspelling recognition system in the preliminary characterization is 60.77%. The average accuracy of 60.77% is lower than the target of 70% for the system accuracy, which was later improved in the improved characterization process that follows.

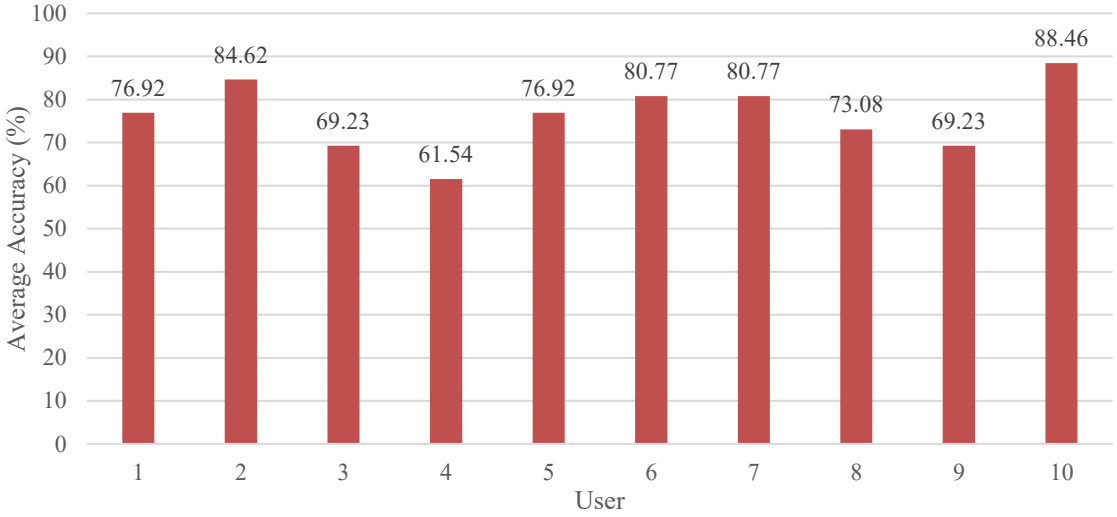


Figure 12. Average accuracy of the fingerspelling recognition system using improved characterization values for ten users.

Once the total average accuracy reached 70% in the improved characterization for the five users tested earlier, the prototype was tested by ten users using the latest fixed threshold values in the improved characterization. The average recognition accuracy of the improved characterization for ten users and all letters is shown in Figs. 12 and 13. It can be observed in Figure 12 that the average recognition accuracy for each user is above 60% in the improved characterization. The highest average recognition accuracy is 88.46% for *User 10,* while the lowest average recognition accuracy is 61.54% for *User 4*. Compared to the preliminary characterization, the prototype shows a significant increment of 19.23% in the lowest user's average recognition accuracy in the improved characterization.

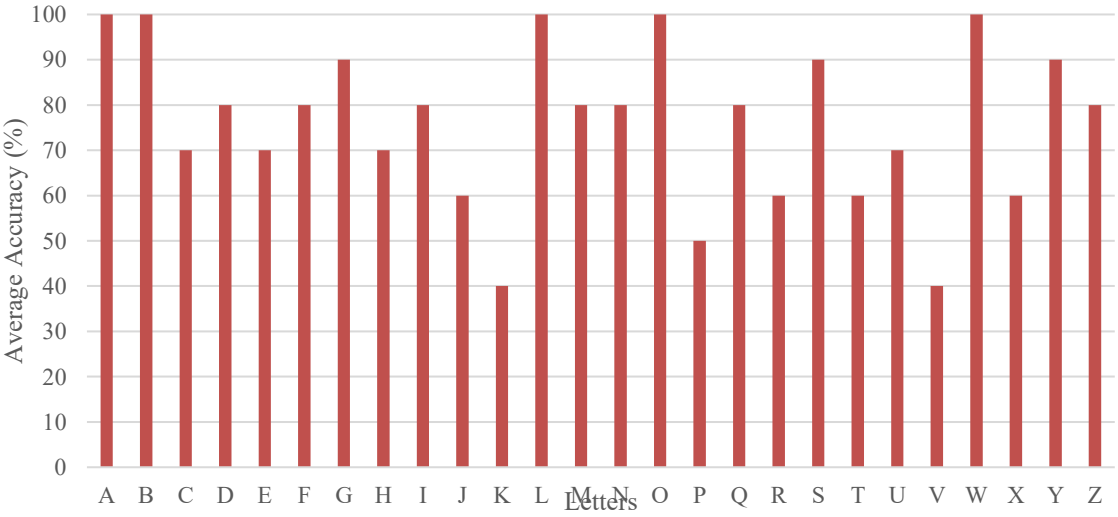


Figure 13. Average accuracy of the fingerspelling recognition system using improved characterization values for each letter detection.

https://doi.org/10.31436/iiumej.v26i3.3633

The average recognition accuracy for all letters using the improved characterization values also recorded significant increment in the lowest average recognition accuracy, which is 40% for the letters 'K' and 'V' as shown in Figure 13. This is because both letters share the same gesture in BIM fingerspelling for index and middle fingers, with slight gesture needed from the thumb for gesturing the letter 'K'. The threshold values in the improved characterization may have mixed up the recognition of these two letters based on the flex sensors' reading, resulting in low average recognition accuracy. Still, only five letters achieved an average recognition accuracy of 100% in the improved characterization, for the letters 'A', 'B', 'L', 'O ', and 'W', compared to nine letters in the preliminary characterization. This makes up for the unrecognized letters in the preliminary characterization, 'J' and 'L', which achieved 60% and 100% average recognition accuracy in each improved characterization. Overall, the total average accuracy for the prototype using the threshold values fixed in the improved characterization process is 76.15%, 15.38% higher than the total average recognition accuracy achieved in the preliminary characterization.

Table 4. Comparison of the proposed work with other reported works

Developed System	System Input	Recognition Processing	Sign Language	System Accuracy	Advantage / Limitation
FingerSpeller [6]	Accelerometers (embedded on five rings) / sensor-based	Hidden Markov Model (HMM) / machine learning	ASL	91.00 %	Promising initiative for real-time communication but requires users to perform data collection for personalized system training.
SpellRing [8]	Microphone and speaker (embedded on a single ring) / sensor-based	Connectionist Temporal Classification (CTC) / machine learning	ASL	89.89 %	Good recognition accuracy but requires the use of a personal computer for recognition processing, which is impractical for real-time applications (bulky hardware).
SHWASI [9]	Flex sensors (attached to a hand glove) / sensor-based	Low-level processing by Arduino UNO	Pre-defined Sign Language	79.00 %	This is a promising approach for real-time communication, but it requires users to learn a new set of predefined finger gestures to use the system, which can be troublesome for some users.
GestureSpeak [7]	Images / vision-based	EfficientNetV2-M and MobileNetV3-L/ deep learning	ASL	98.42 %	High recognition accuracy, but not yet deployed to an embedded hardware platform for real-time communication and use.
Proposed Work	Flex sensors (attached to a hand glove) / sensor-based	Conventional if-else algorithm by Arduino MEGA 2560	BIM	76.15 %	Ready to be used in real-time communication with promising recognition accuracy.

https://doi.org/10.31436/iiumej.v26i3.3633

The average recognition accuracy of 76.15% shows that the prototype presented in this paper has the potential to be further refined for improvement as a real-time fingerspelling recognition system. Although the average recognition accuracy is the lowest compared to the other sensor-based systems mentioned, such as FingerSpeller (91%), SpellRing (89.89%), and SHWASI (79%), none of the mentioned systems utilize BIM fingerspelling. Besides, the mentioned systems mostly needed extra effort from the user for input training, either to enlarge the training datasets (FingerSpeller and SpellRing) or to get familiar with the pre-defined set of finger gestures customized for the system (SHWASI). This reduces the system's flexibility in being directly used in real-time communication. Still, the constant effort in the development approach for all the mentioned systems shows that researchers are concerned about the system's enhancement and the possibility of further benefiting the hearing and speech impairment community.

Table 4 compares the proposed work in this paper with the other promising fingerspelling recognition systems that other researchers have reported. It can be observed that while all developed systems are in the experimental and development phase, the proposed work in this paper is the closest initiative that can be utilized in real-time communication for a BIM fingerspelling recognition system.

5. CONCLUSION

This paper shares the flex sensor resistance value modeling for a fingerspelling recognition system using two main processes: preliminary characterization and improved characterization. The flex sensor resistance value was converted into its corresponding ADC readings, which were utilized for the threshold values in the initial characterization and improved characterization processes. In the preliminary characterization, user threshold values recorded for each flex sensor's resistance values produced for all fingerspelled letters were used in the fingerspelling recognition processing. The total average recognition accuracy when five users tested the system in this preliminary characterization process was 60.77%, which is lower than the average accuracy intended as the proof-of-concept of this prototype, which is 70%. The prototype was constructed using flex sensors and an accelerometer as inputs, Arduino MEGA 2560 as processing platform, and LEDs, buzzer, and OLED display as outputs.

In the improved characterization process, flex sensors' resistance values recorded earlier from the five users were analyzed to adjust the threshold values in the fingerspelling recognition processing. The fingerspelling recognition processing used the conventional if-else algorithm to map the hand gestures into their fingerspelled letters via the threshold values configured. The five users tested the prototype repeatedly using the improved threshold values until it achieved at least 70% recognition accuracy. Once the average recognition accuracy reached the target of 70%, the latest threshold values were fixed in the fingerspelling recognition processing algorithm. The total average accuracy of the system using the latest improved characterization threshold value is 76.15%, which is 6.15% higher than the targeted 70% value.

Still, there is room for improvement in this work, as the accuracy of 76.15% achieved is 23.85% less than the perfect accuracy of 100%. This might be due to the hand glove of this prototype, which is standard to one size and affected the system's accuracy, given the various sizes of users' hands during the demonstration. As the flex sensor was fitted at the same position along the length of the fingers on the hand glove, the bending of the flex sensors for different hand sizes imposed different resistance values due to the flex sensor's placement. The hand glove may fit nicely for one user but may be too loose for another due to different hand sizes. As such, the threshold values fixed for the system might work perfectly for one user

during the fingerspelling recognition system, but not for the following user due to variation in the sensor's placement, which had already been fitted along the length of the fingers on the same hand glove. Future work in this direction can focus on a customized hand glove to be used for different hand sizes, which can be grouped into either male or female users.

Other potential future work includes consideration for an adjustable flex sensor placement on the hand glove, instead of preparing several separate hand gloves for individual sizes. The critical thing in this adjustable placement is to ensure that the flex sensor is fixed at the knuckles or joints of the fingers, where the bend of the fingers can be accurately measured. In this adjustment, the system can utilize the two available off-the-shelf (OTS) flex sensors, lengths that are either 2.2 inches or 4.5 inches, instead of only 4.5-inch flex sensors as utilized in this proposed work.

Furthermore, the prototype developed in this paper was tested by normal hearing individuals who are not frequent sign language users. This is another limitation of this work, where future work will include prototype demonstrations by users with hearing and speech impairments who are frequent sign language users. Suppose frequent fingerspelling users tested the prototype. In that case, the recognition accuracy achieved by the system is more focused on the targeted group of users, and the improvements made will directly cater to the needs of these users.

ACKNOWLEDGEMENT

Deepest appreciation to Universiti Teknikal Malaysia Melaka (UTeM) for all the support given during the completion of this research work. The prototype shown in this paper had also been showcased in the 2nd edition of the International Research & Information Science Expo 2025 (iRISE25) and awarded the gold medal.

REFERENCES

[1] Advancing hearing health in Malaysia: a harmonious call to action — Olivier Chupin. Available: https://codeblue.galencentre.org/2023/10/advancing-hearing-health-in-malaysia-a-harmonious-call-to-action-olivier-chupin/

[2] Raising awareness for the deaf community in the country. Available: https://www.bernama.com/en/thoughts/news.php?id=2251189

[3] Hilario-Acuapan G, Ordaz-Hernández K, Castelán M, Lopez-Juarez I. (2025) Toward a recognition system for Mexican sign language: arm movement detection. Sensors, 25:3636. doi:10.3390/s25123636

[4] Herrick YHL, Norhanifah M. (2022) BIM sign language translator using machine learning (TensorFlow). Journal of Soft Computing and Data Mining, 3(1): 68-77.

[5] Nurrahma N, Rahadian Y, Ary SP. (2021) Indonesian sign language fingerspelling recognition using vision-based machine learning. In Proceedings of the 2021 International Conference on Intelligent Cybernetics Technology and Applications (ICICyTA): 1-2 December 2021; virtually. doi:10.1109/ICICyTA53712.2021.9689176

[6] David M, Zikang L, Tan G, Jon W, Jocelyn H, Bill N, Hyeokhyen K, Thomas P, Thad S. (2023) FingerSpeller: camera-free text entry using smart rings for American sign language fingerspelling recognition. In Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '23): 22-25 October 2023; New York. doi:10.1145/3597638.3614491

[7] Wei HY, Sin TL, Rui C, Ashwaq MSQ. (2024) GestureSpeak: An efficient real-time finger spelling recognition model. In Proceedings of the 2024 International Conference on Computing Innovation, Intelligence, Technologies and Education (CIITE): 5-6 September 2024; Malaysia. pp 1-6.

[8] Hyunchul L, Nam AD, Dylan L, Tianhong CY, Jane L, Franklin ML, Yiqi J, Yan M, Xiaojun B, Francois G, Cheng Z. (2025) SpellRing: recognizing continuous fingerspelling in American sign language using a ring. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems: 26 April 2025 – 1 May 2025; Yokohama. pp 1-17.

[9] Mithun E, Dharun KS, Abirami A, Mathiarasun RJ, Harikumar ME. (2024) (SHWASI) Smart hand wearable aid for speech impaired: sign language communication using flex sensor-based finger spelling. In Proceedings of the 2024 IEEE 21st India Council International Conference (INDICON): 19-21 December 2024; India. pp 1-6.

[10] Marie A, Hamid RM, Ivan G. (2023) A survey on sign language literature. Machine Learning with Applications, 14: 100504. doi:10.1016/j.mlwa.2023.100504

[11] Yuvraj G, Riya A, Deepak S, Prashant KG. (2021) Sign language translation systems for hearing/speech impaired people: a review. In Proceedings of the 2021 International Conference on Innovative Practices in Technology and Management (ICIPTM): 17-19 February 2021; India. pp 10-14.

[12] Bahasa Isyarat Malaysia (BIM): Abjad. Available: https://www.bimsignbank.org/groups/general/alphabets

[13] Razieh R, Kourosh K, Sergio E. (2021) Sign language recognition: a deep survey. Expert Systems with Applications, 164: 113794.

[14] Md. AAF, Farhan FA, Mosabber UA. (2021) Sensor dataglove for real-time static and dynamic hand gesture recognition. In Proceedings of the 2021 Joint 10th International Conference on Informatics, Electronics & Vision (ICIEV) and 2021 5th International Conference on Imaging, Vision & Pattern Recognition (icIVPR): 16-20 August 2021; Japan. doi:10.1109/ICIEVicIVPR52578.2021.9564226.

[15] Zinah RS, Zurinahni Z, Zaidan BB, Alamoodi AH. (2022) A systematic review on systems-based sensory gloves for sign language pattern recognition: an update from 2017 to 2022. IEEE Access, 10:123358-123377.

[16] Santosh KS, Ravi G, Sunil KKN, Nataraja N, Rajendra PP, Sarala T. (2021) Glove based deaf-dumb sign language interpreter. In Proceedings of the 2021 6th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT): 27-28 August 2021; India. pp 947-950.

[17] Lu C, Amino S, Jing L. (2023) Data glove with bending sensor and inertial sensor based on weighted DTW fusion for sign language recognition. Electronics, 12(3):613. doi:10.3390/electronics12030613

https://doi.org/10.31436/iiumej.v26i3.3634

PROPOSED CONVBILSTM-NET MODEL FOR ENHANCING EARTHQUAKE PREDICTION PERFORMANCE USING SPATIOTEMPORAL FEATURES

ARI FADLI[1,3], TRI KUNTORO PRIAMBODO[1*], AGFIANTO EKO PUTRA[1], WIWIT SURYANTO[2]

[1]*Department of Computer Science and Electronics, Universitas Gadjah Mada, Yogyakarta, Indonesia*
[2]*Department of Physics, Universitas Gadjah Mada, Yogyakarta, Indonesia*
[3]*Electrical Engineering, Universitas Jenderal Soedirman, Purwokerto, Indonesia*

**Corresponding author: mastri@ugm.ac.id*

(Received: 3 March 2025; Accepted: 15 July 2025; Published online: 9 September 2025)

ABSTRACT: Accurate earthquake prediction remains a significant challenge due to the complex spatiotemporal dependencies inherent in seismic events. To address this issue, the present study proposes ConvBiLSTM-Net. This hybrid deep learning model combines Convolutional Neural Networks (CNNs) for spatial feature extraction with Bidirectional Long Short-Term Memory (BiLSTM) networks for temporal sequence modeling. The model integrates historical earthquake data with spatial information in the form of fault density (FD), derived using Kernel Density Estimation (KDE). The KDE bandwidth is optimized using the Bivariate Local Indicator of Spatial Association (LISA) method to enhance spatial adaptivity. The dataset comprises earthquake records from the USGS catalog (1974–2023) and active fault data compiled in the 2017 Indonesian Earthquake Source and Hazard Map, published by the National Earthquake Study Center (PuSGeN). ConvBiLSTM-Net is evaluated under short-term and medium-term prediction scenarios, targeting earthquake magnitude, depth, and epicenter coordinates (latitude and longitude), using standard performance metrics such as accuracy, F1 score, root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2). In the short-term scenario, the model achieves average improvements of 9.31% in R^2, 3.41% in accuracy, and 6.06% in F1 score, while reducing RMSE by 10.63% and MAE by 12.40% across magnitude, depth, and latitude predictions. For longitude, R^2, accuracy, and F1 score also improve by 10.88%, 11.76%, and 17.54%, respectively, although RMSE and MAE increase by 13.09% and 20.74%, indicating a trade-off between enhanced pattern recognition and higher absolute error. Under the medium-term scenario, the model demonstrates average improvements of 7.49% in R^2, 3.39% in accuracy, and 7.06% in F1 score, while reducing RMSE and MAE by 6.22% and 17.72%, respectively, for magnitude, depth, and latitude predictions. For longitude, R^2, accuracy, and F1 score improve by 12.50%, 2.48%, and 1.60%, respectively, though RMSE and MAE increase by 37.31% and 37.01%, again highlighting a trade-off between better pattern recognition and increased absolute error in this dimension. These findings demonstrate that ConvBiLSTM-Net, engineered to integrate spatial and temporal features, is a robust and adaptive architecture for enhancing earthquake prediction performance. Its spatiotemporal modeling approach yields consistently high accuracy and stability across forecasting horizons, particularly in predicting earthquake epicenters. Despite minor trade-offs in absolute error for longitude predictions, the overall performance improvements affirm its potential as a reliable tool for seismic hazard assessment and disaster risk mitigation.

ABSTRAK: Ramalan gempa bumi yang tepat kekal sebagai satu cabaran utama disebabkan oleh kebergantungan spatiotemporal yang kompleks dalam kejadian seismik. Bagi menangani isu ini, kajian ini mencadangkan ConvBiLSTM-Net, iaitu sebuah model hibrid pembelajaran mendalam yang menggabungkan Rangkaian Neural Konvolusional (CNN) dan Memori

Jangka Pendek Dwi Arah (BiLSTM), bagi tujuan pengekstrakan ciri spatial dan pemodelan jujukan temporal, masing-masing. Model ini menggabungkan data sejarah gempa bumi dengan maklumat spatial dalam bentuk ketumpatan sesar (fault density, FD), yang diperoleh melalui Kaedah Anggaran Ketumpatan Kernel (Kernel Density Estimation, KDE). Lebar jalur KDE dioptimumkan menggunakan kaedah Bivariate Local Indicator of Spatial Association (LISA) bagi meningkatkan kepekaan spatial. Set data kajian merangkumi rekod gempa bumi daripada katalog USGS (1974–2023) serta data sesar aktif yang disusun dalam Peta Sumber dan Bahaya Gempa Indonesia 2017, terbitan Pusat Kajian Gempa Nasional (PuSGeN). Model ConvBiLSTM-Net ini dinilai dalam dua senario ramalan—jangka pendek dan jangka sederhana—bagi parameter magnitud, kedalaman, serta koordinat pusat gempa (latitud dan longitud), dengan menggunakan metrik standard seperti ketepatan, skor F1, RMSE, MAE dan pekali penentuan (R^2). Malalui senario jangka pendek, model mencatatkan purata peningkatan sebanyak 9.31% pada R^2, 3.41% pada ketepatan, dan 6.06% pada skor F1; serta pengurangan RMSE sebanyak 10.63% dan MAE sebanyak 12.40% merentas ramalan magnitud, kedalaman, dan latitud. Bagi dimensi longitud, R^2, ketepatan, dan skor F1 turut meningkat sebanyak 10.88%, 11.76%, dan 17.54% masing-masing; namun begitu, RMSE dan MAE meningkat sebanyak 13.09% dan 20.74%, menunjukkan kompromi antara pengecaman corak yang lebih baik dan ralat mutlak yang lebih tinggi. Manakala senario jangka sederhana, model mencatatkan purata peningkatan sebanyak 7.49% pada R^2, 3.39% pada ketepatan, dan 7.06% pada skor F1; serta pengurangan RMSE sebanyak 6.22% dan MAE sebanyak 17.72% merentas tugas ramalan magnitud, kedalaman, dan latitud. Bagi dimensi longitud, peningkatan masing-masing dicatatkan pada R^2 (12.50%), ketepatan (2.48%), dan skor F1 (1.60%), tetapi RMSE dan MAE meningkat secara ketara sebanyak 37.31% dan 37.01%, menunjukkan kompromi antara pengecaman corak yang lebih kukuh dan ralat mutlak yang lebih besar pada dimensi ini. Dapatan kajian ini membuktikan bahawa ConvBiLSTM-Net, yang direka bentuk bagi menggabungkan ciri-ciri spatial dan temporal, merupakan satu seni bina model yang teguh dan adaptif dalam meningkatkan prestasi ramalan gempa bumi. Pemodelan spatiotemporal bersepadu yang digunakan menghasilkan tahap ketepatan dan kestabilan yang tinggi secara konsisten merentasi pelbagai ufuk ramalan, terutamanya dalam menentukan lokasi pusat gempa bumi. Walaupun terdapat sedikit kekurangan dalam nilai ralat mutlak bagi ramalan longitud, peningkatan prestasi secara keseluruhan mengesahkan nilainya sebagai alat yang boleh dipercayai dalam penilaian bahaya seismik dan pengurangan risiko bencana.

KEYWORDS: *Deep learning, earthquake prediction, kernel density estimation (KDE), spatial–temporal features.*

1. INTRODUCTION

Earthquakes are among the most destructive natural disasters, occurring without warning and resulting in significant human and economic losses [1]. They can trigger secondary hazards such as fires, infrastructure collapse, and tsunamis, further exacerbating their impact on lives and property. Since 1998, more than 750,000 people have died due to earthquakes, primarily in seismically active countries such as Japan, China, and Indonesia [2]. Based on their causes, earthquakes are generally categorized into three types: tectonic earthquakes (caused by plate movements), volcanic earthquakes (resulting from volcanic activity), and impact earthquakes (caused by collisions with extraterrestrial objects such as meteorites) [3]. Several significant events have resulted in catastrophic damage, including the 2011 Tōhoku earthquake in Japan (magnitude 9.1, approximately 20,000 deaths, and $360 billion in losses), the 2015 Nepal earthquake (magnitude 7.8, 8,964 deaths, and $10 billion in losses), and the 2023 Turkey–Syria earthquake (magnitude 7.5, 45,000 deaths, and $100 billion in losses) [4]. These disasters underscore the critical need for global cooperation in disaster preparedness, response, and recovery to mitigate future risks.

Earthquake prediction is typically classified into three categories based on temporal scale: long-term prediction (ranging from hundreds to thousands of years), medium-term prediction (spanning decades to a few years), and short-term prediction (ranging from a few days to one week) [5-7]. This classification has served as a conceptual framework for developing predictive methodologies across various disciplines, including mathematics, statistics, and artificial intelligence (AI) [8]. While statistical and mathematical models have shown consistent utility in earthquake forecasting [9,10], their limitations in capturing seismic phenomena‘ highly nonlinear and spatiotemporal nature remain evident.

Recent advances in AI, particularly in machine learning (ML) and deep learning (DL), have introduced a paradigm shift in predictive seismology. These data-driven approaches excel in extracting latent patterns from complex, high-dimensional seismic datasets without relying on handcrafted features. A comprehensive meta-analysis of 84 scientific publications highlights that AI-based methods outperform traditional prediction accuracy and adaptability models, ranging from rule-based fuzzy inference systems to modern deep neural networks. Furthermore, these approaches offer scalable, region-specific solutions for real-time earthquake forecasting. This growing body of evidence reinforces the transformative role of AI in bridging critical gaps in seismic risk assessment and early warning systems, particularly for short-term prediction scenarios [11].

Consequently, AI-based techniques are increasingly recognized as indispensable tools for enhancing modern earthquake prediction systems’ accuracy, responsiveness, and reliability. The broader success of AI, including ML and DL, in predictive analytics further underscores their relevance. Enabled by rapid advancements in computational power, these methods have demonstrated exceptional capabilities in identifying complex patterns within large-scale datasets, making them particularly effective for addressing nonlinear and dynamic forecasting challenges. Their influence is especially notable in information science and technology, where they have been successfully applied across various domains [12-14].

For example, Huang et al. [15] employed an ensemble approach for wind speed prediction; Peng et al. [16] combined Random Forest with Long Short-Term Memory (LSTM) networks to forecast tourist arrivals; Li and Wu [17] utilized clustering techniques for market style prediction; and Wang et al. [18] applied ensemble methods to model customer churn. Moreover, AI- and DL-based approaches have demonstrated strong capabilities in anomaly detection, further validating their applicability across diverse predictive modeling tasks. These cross-domain applications highlight the potential of AI-driven frameworks to address the complex spatiotemporal characteristics inherent in earthquake prediction.

ML and DL methodologies have been extensively adopted in earthquake prediction research, commonly formulated as classification or regression problems. Classification-based approaches typically frame earthquake prediction as a binary decision task, determining whether an earthquake exceeding a predefined magnitude threshold will likely occur within a specific time window. For instance, Asim et al. employed various ML classifiers to predict the occurrence of earthquakes with magnitudes ≥ 5.5 in the Hindukush region within a one-month forecasting window, utilizing seismic indicators derived from historical earthquake catalogs [19]. In a subsequent study, Asim et al. [20] implemented Support Vector Regression (SVR) and Hybrid Neural Networks (HNN), optimized via Enhanced Particle Swarm Optimization (EPSO), to predict earthquakes with magnitudes greater than 5.0 in regions including the Hindu Kush, Chile, and Southern California. Similarly, Yaghmaei-Sabegh and Tsang [21] developed an Artificial Neural Network (ANN)-based site classification model using horizontal-to-vertical spectral ratio curves. Additional studies have integrated seismicity analysis with ML

algorithms, such as Support Vector Machines (SVM), Random Forest (RF), and ANN, to forecast earthquake occurrences in regions like Cyprus [22].

Regression-based models, in contrast, aim to predict specific earthquake parameters such as magnitude, timing, and location. Jain et al. developed predictive models leveraging location and depth features from United States Geological Survey (USGS) data, applying a range of ML algorithms including RF, Multi-Layer Perceptron (MLP), and SVR. Their findings revealed that the MLP model achieved the lowest root mean square error (RMSE) among all tested models, indicating superior performance in earthquake location prediction [23]. Similarly, Gonzalez et al. proposed using Recurrent Neural Networks (RNNs), specifically Gated Recurrent Units (GRUs), to model the spatiotemporal dynamics of seismic activity in Italy. Their model was trained on seismic data from 1995 to 2018 and could estimate the location, timing, magnitude, and depth of earthquakes moments before they occurred [24]. These findings highlight the efficacy of regression-based approaches, particularly deep learning architectures such as MLP and GRU, in capturing complex spatial and temporal relationships essential for accurate earthquake parameter prediction.

DL techniques have substantially improved earthquake forecasting because they can model intricate temporal dependencies using deep neural architectures. These models, which consist of multiple hidden layers and interconnected neurons, demonstrate enhanced generalization capabilities compared to shallow networks [25]. Wang et al. applied Long Short-Term Memory (LSTM) networks to capture spatiotemporal correlations among earthquake events across various regions in China. Their model included a dropout layer to mitigate overfitting, employed the Softmax function for activation, and was optimized using the RMSprop algorithm. The approach yielded superior predictive performance relative to traditional techniques [8]. Similarly, Cai et al. [26] proposed an LSTM-based prediction model for automated anomaly detection in earthquake precursor data, drawing on multi-disciplinary datasets. Their results indicated that LSTM networks effectively captured temporal patterns in normal behavior across diverse precursor sources, without requiring extensive domain-specific preprocessing. These outcomes suggest that LSTM networks possess strong potential for modeling complex temporal dynamics in seismic forecasting.

Fabregas et al. developed a system that integrates a Rule-Based Algorithm with LSTM networks to estimate earthquake frequency, maximum magnitude, and average depth over one year [27]. Berhich et al. [28] proposed a location-aware earthquake prediction approach using recurrent neural networks, including LSTM, Gated Recurrent Units (GRU), and their hybrid configurations. They applied K-Means clustering based on geographic features to segment seismic data from Morocco, Japan, and Turkey, effectively capturing region-specific seismic patterns. The results demonstrated strong predictive performance, particularly for high-magnitude events, with improved mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE) compared to previous studies.

Kail et al. [29] introduced a hybrid neural network combining Convolutional Neural Networks (CNNs) for spatial feature extraction with LSTM networks for temporal modeling. This integrated architecture significantly enhanced mid-term earthquake prediction performance, reducing false alarms by approximately 81% while maintaining comparable sensitivity in detecting actual events. Similarly, Bhandarkar et al. [30] empirically demonstrated that LSTM models significantly outperform feedforward neural networks (FFNN) in modeling earthquake trends, achieving a 59% improvement in the coefficient of determination (R^2).

https://doi.org/10.31436/iiumej.v26i3.3634

Collectively, these studies highlight the effectiveness of hybrid and recurrent deep learning models in earthquake prediction. By integrating CNNs for spatial feature extraction with LSTM or GRU networks for temporal sequence modeling, these approaches have achieved notable improvements in predictive accuracy, including reduced error rates, enhanced sensitivity to region-specific seismic characteristics, and substantially lower false alarm rates, thereby offering a robust framework for reliable seismic forecasting.

Recent research has further advanced hybrid deep learning models to improve the accuracy and generalizability of earthquake prediction. Kavianpour et al. [31] developed a hybrid model incorporating Attention Mechanisms, CNN, BiLSTM, and Zero-Order Hold (ZOH) preprocessing to forecast earthquake magnitude and frequency across seismic regions in China. This model achieved superior performance, with an RMSE of 0.074, MAE of 0.076, and an R^2 of 0.906. In a related study, Kavianpour et al. [32] designed a CNN–BiLSTM model to predict monthly mean earthquake magnitudes in Japan. Building upon this foundation, the present study introduces an enhanced architecture, HC_BiLSTM, with several refinements to improve predictive accuracy. The model achieved the lowest RMSEs of 0.1170 and 0.0982 across two study areas.

Geng et al. [33] proposed deep temporal convolutional networks (DCTCNN and CNN-LSTM) to address long-term dependencies in seismic time series data. Their models outperformed traditional methods such as ARIMA, Support Vector Regression (SVR), Random Forest (RF), and baseline deep learning approaches. Specifically, DCTCNN achieved an RMSE of 0.0061 and an MAE of 0.070, corresponding to over 96.9% improvement in RMSE and 55.97% in MAE.

These recent advancements underscore the effectiveness of hybrid deep learning architectures in earthquake prediction. Models that combine convolutional layers (e.g., CNN or Temporal Convolutional Networks [TCN]) for spatial feature extraction with recurrent layers (e.g., LSTM or BiLSTM) to capture temporal dependencies, often enhanced by Attention Mechanisms or ZOH preprocessing, consistently outperform both traditional statistical methods and standalone deep learning models.

Despite recent advances, most studies have focused exclusively on temporal patterns, without explicitly incorporating spatial information into the modeling process [34]. However, empirical evidence indicates that destructive earthquakes predominantly occur near active fault zones [35-37], highlighting the importance of spatial features such as Fault Density (FD). FD represents the level of tectonic fragmentation in the Earth's crust and can be estimated using the spatial distribution of earthquake events around fault lines [34,38].

To address this research gap, the present study proposes a novel hybrid deep learning model called ConvBiLSTM-Net, which integrates spatial and temporal dependencies. The architecture combines Convolutional Neural Networks (CNNs) for spatial feature extraction with Bidirectional Long Short-Term Memory (BiLSTM) networks for temporal sequence modeling. Spatial features include fault density (FD), computed via kernel density estimation (KDE), with bandwidth optimized using a bivariate Local Indicator of Spatial Association (LISA), in contrast to previous studies that employed Bivariate Moran's Index for bandwidth selection [34,38]. This study adopts bivariate LISA to capture localized spatial autocorrelation in fault distribution better, as it enables the detection of spatial clusters and outliers that may be masked by global measures such as Bivariate Moran's Index [39]. Temporal features are derived from historical earthquake sequences. The remainder of this paper is structured as follows: Section 2 outlines the methodology, Section 3 presents the results and discussion, and Section 4 concludes the study.

https://doi.org/10.31436/iiumej.v26i3.3634

2. METHODOLOGY

This section outlines the methodology for developing the proposed earthquake prediction model based on deep learning. The model was implemented in Python 3.9, utilizing TensorFlow 2.15 (Keras API) for deep learning, scikit-learn 1.4 for data preprocessing, GeoPandas 0.14 for spatial data manipulation, and GDAL 3.8 for raster data processing. All experiments were conducted on a local computing environment equipped with an NVIDIA GeForce RTX 3050 Laptop GPU and 16 GB of RAM. This provided adequate computational efficiency and ensured reproducibility throughout the study. The model is designed to predict earthquake magnitude, depth, and location. A comprehensive earthquake catalog was employed to enhance the accuracy and reliability of the prediction outcomes.

2.1. Architecture of the ConvBiLSTM-Net Model

The deep learning model developed in this study to predict earthquake magnitude, depth, and epicentral location is termed ConvBiLSTM-Net. The overall architecture of the model is illustrated in Fig. 1. It integrates Convolutional Neural Networks (CNNs) for spatial feature extraction, Bidirectional Long Short-Term Memory (BiLSTM) layers for temporal sequence learning, and fully connected Dense layers for final prediction. This hybrid architecture is designed to effectively capture the complex spatiotemporal dependencies inherent in earthquake catalog data.

The network begins with an input layer that receives feature vectors comprising both temporal indicators (e.g., statistical seismic features) and spatial attributes (e.g., fault density). These inputs are processed through a convolutional block consisting of a convolutional layer, a batch normalization layer, and a max pooling layer.

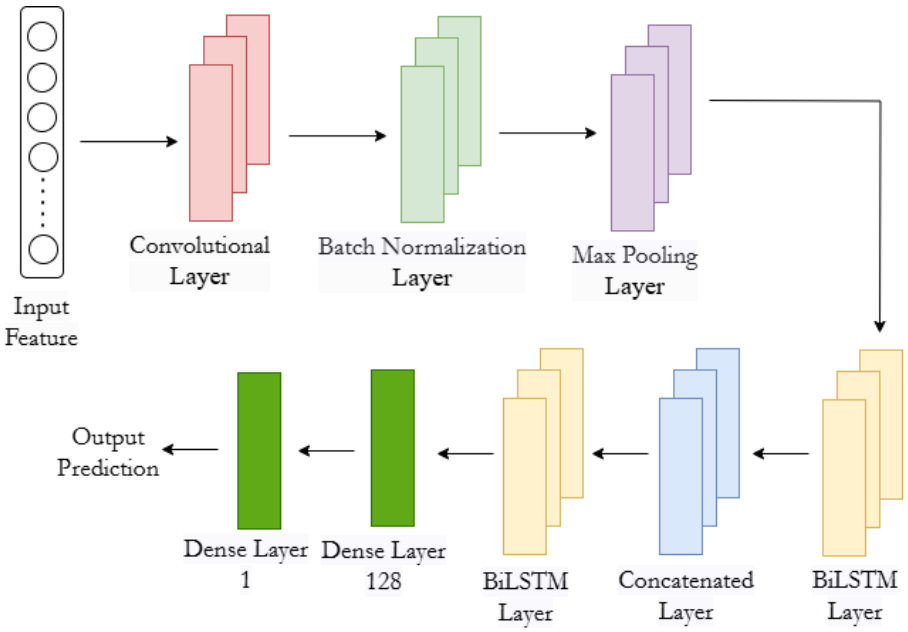


Figure 1. The structure of the ConvBiLSTM-Net

The convolutional layer applies multiple filters to extract local feature patterns. The output at position (i, j) in the l^{th} layer is computed as in Equation (1) [32].

$$y_{i,j}^{l} = ReLU\left(b_j^l + \sum_{m=1}^{M} w_{m,j}^{l} * x_{i+m-1,j}\right) \quad (1)$$

https://doi.org/10.31436/iiumej.v26i3.3634

Here, * denotes a 1D convolution operation, b_j^l is the bias term, ω represents the filter weights, and m indexes the kernel position. The activation function used is ReLU, defined as $RELU\ (x) = \max(0, x)$, which enables efficient convergence and mitigates the vanishing gradient problem. A max pooling layer follows to prevent overfitting and reduce spatial dimensionality by retaining the most prominent features within a pooling window. The operation is defined in Equation (2) [40]:

$$p_{i,j}^l = \max\left(y_{i.T+r,j}^{l-1}\right) \tag{2}$$

The model also incorporates a Batch Normalization (BN) layer to stabilize training and accelerate convergence. BN normalizes the inputs using mini-batch statistics and applies a learned scaling and bias transformation, as follows where x_i and y_i denote the input and output for the i_{th} sample, respectively, γ and β are learnable parameters, and ε is a small constant added for numerical stability [41].

$$\mu = \frac{1}{N_{batch}} \sum_{i=1}^{N} x_i \tag{3}$$

$$\sigma^2 = \frac{1}{N_{batch}} \sum_{i=1}^{N} (x_i - \mu)^2 \tag{4}$$

$$\widehat{x_i} = \frac{x_i - \mu}{\sqrt{\sigma^2 - \varepsilon}} \tag{5}$$

$$\mu y_i = \gamma x_i + \beta \tag{6}$$

Following the CNN block, a flattening layer reshapes the output into a one-dimensional array suitable for input into the BiLSTM layer, which is designed to capture sequential patterns in both forward and backward time directions. Compared to conventional RNNs, BiLSTM can retain long-term dependencies and reduce gradient degradation, making it well-suited for modeling earthquake sequences.

In the LSTM architecture, memory updates are controlled by three gates: the forget gate (f_t), which discards irrelevant information; the input gate (i_t), which updates the memory state; and the output gate (o_t), which determines the output based on the current memory state. These gates regulate information flow through the memory cell using sigmoid activations. The detailed gate computations, adapted from [32], are presented in the following equations:

$$f_t = \sigma\left(W_f . [h_{t-1}, p_t] + b_f\right) \tag{7}$$

$$i_t = \sigma(W_i . [h_{t-1}, p_t] + b_i) \tag{8}$$

$$\widehat{C_t} = tanh(W_C . [h_{t-1}, p_t] + b_C) \tag{9}$$

$$o_t = \sigma(W_o . [h_{t-1}, p_t] + b_o) \tag{10}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \widehat{C_t} \tag{11}$$

Here, h_{t-1} denotes the hidden state from the previous time step, p_i is the current input, C_t is the cell state, and σ represents the sigmoid activation function.

2.2. Methodological Framework

Figure 2 illustrates the methodological framework employed in this study. The process begins with acquiring earthquake data from the United States Geological Survey (USGS), comprising 37,316 earthquake events recorded between 1974 and 2023. The study focuses on the Indonesian region, with the spatial domain defined within the geographic extent of 95°–141°E longitude and 11°S–6°N latitude, encompassing the entire national territory.

https://doi.org/10.31436/iiumej.v26i3.3634

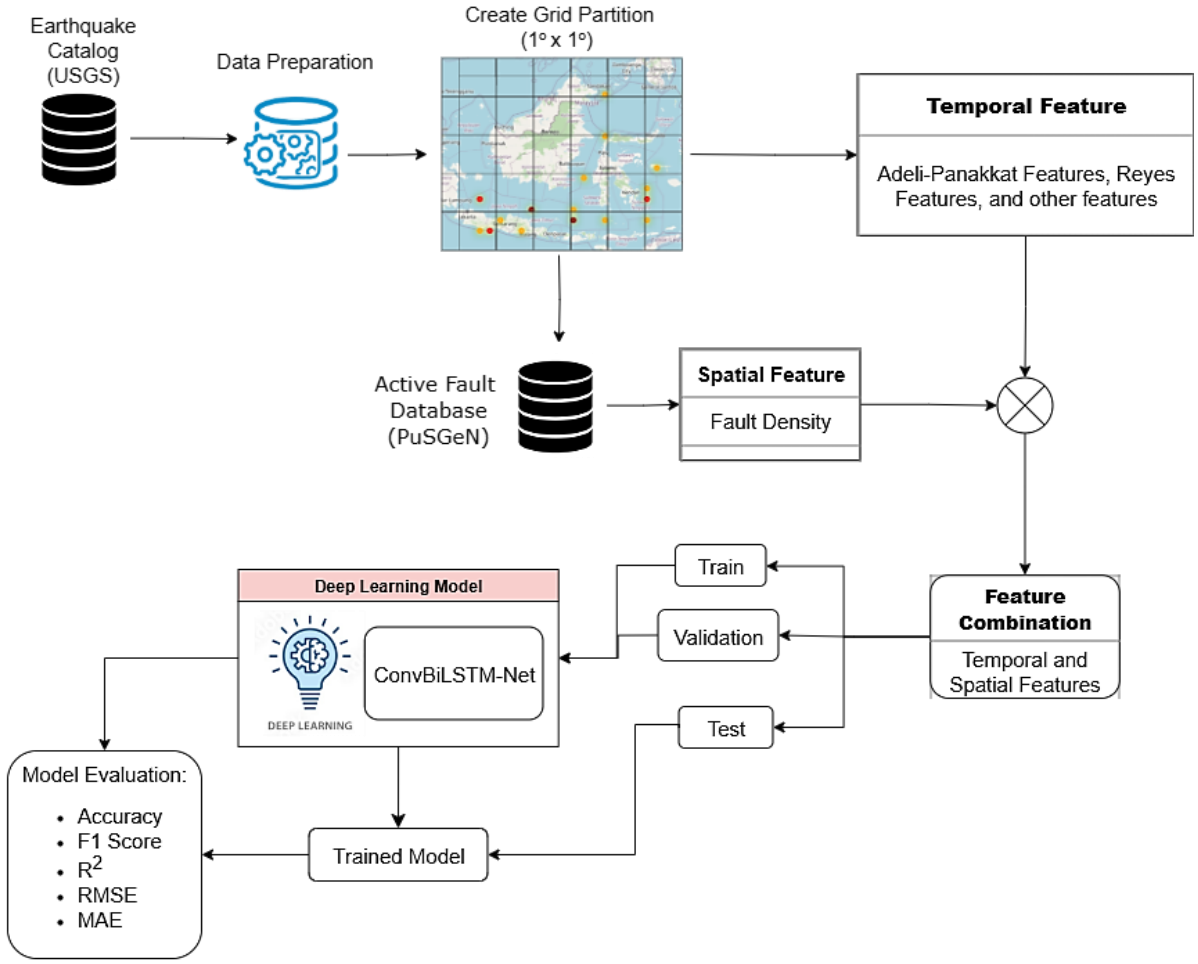


Figure 2. Methodological framework for earthquake prediction

Following data acquisition, preprocessing steps are undertaken to remove null values, duplicates, and irrelevant entries. The study area is subsequently partitioned into a grid of 1°×1° cells to facilitate spatial structuring. Each grid cell is enriched with spatial features derived from active fault data compiled in the 2017 Indonesian Earthquake Source and Hazard Map, published by the National Earthquake Study Center (PuSGeN), which includes mapped segments of active faults across the Indonesian region [42].

To characterize the spatial dimension of seismic activity, fault density (FD) is calculated using a Gaussian Kernel Density Estimation (KDE) approach. To optimize spatial representation, this study introduces a novel strategy for selecting the KDE bandwidth using the Bivariate Local Indicator of Spatial Association (LISA). This method identifies the bandwidth that yields the highest number of statistically significant local clusters (e.g., high–high, low–low), thereby ensuring spatial coherence between seismic patterns and geological fault distributions.

https://doi.org/10.31436/iiumej.v26i3.3634

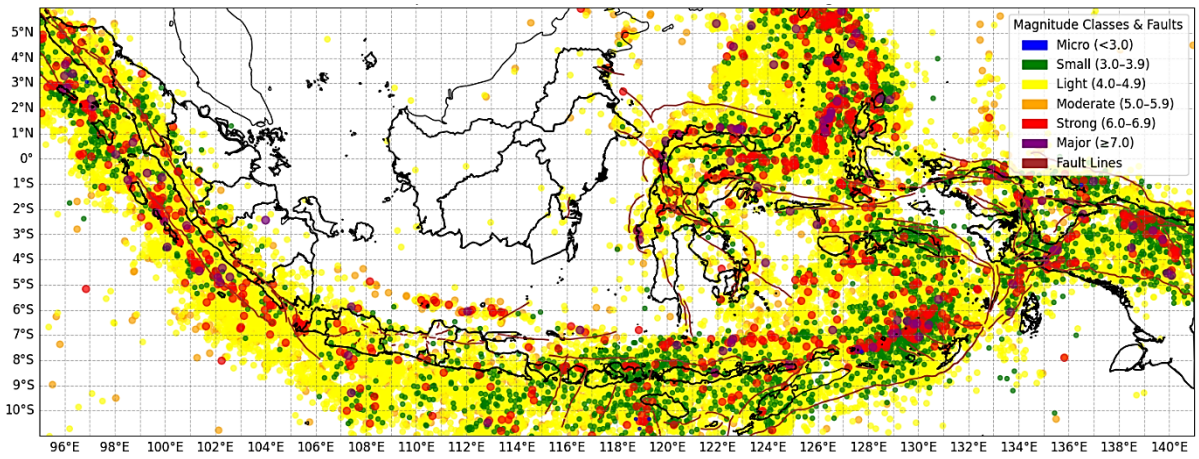


Figure 3. Spatial distribution of earthquake events in Indonesia (1974–2023), overlaid with active fault lines and classified by magnitude ranges

In parallel, the temporal dimension is captured using a comprehensive set of features derived from three established sources. First, nine statistical indicators adapted from Panakkat and Adeli [43] are employed, as presented in Table 1, including seismic productivity (a- and b-values), energy release, and inter-event times based on the Gutenberg–Richter law.

Table 1. Adeli–Panakkat Features

#	Feature	Description
1	b-value	Slope of the magnitude–log frequency plot (Gutenberg–Richter law).
2	a-value	Intercept of the Gutenberg–Richter law, representing seismic productivity.
3	H	Sum of the mean squared deviations from the regression line based on the Gutenberg–Richter law.
4	ΔM	Difference between the largest observed magnitude and the largest expected value according to the Gutenberg–Richter law.
5	T	Elapsed time between the most recent *N* earthquake events.
6	M	Average time between major seismic (characteristic) events among the last *N* events.
7	C	Coefficient of variation of the average time between characteristic events.
8	$dE^{1/2}$	The rate of the square root of the energy released.
9	M_{Mean}	Mean magnitude of the last *N* earthquake events.

In addition, seven temporal evolution metrics proposed by Reyes et al. [44] are incorporated, as summarized in Table 2. These features include successive changes in b-values and the probability of large-magnitude events. Furthermore, basic geophysical attributes, such as magnitude, depth, elapsed time, energy, and spatial coordinates, are extracted directly from the earthquake catalog, as listed in Table 3.

Table 2. Reyes Features

#	Feature	Description
1	X1	Increment of the *b*-value between events *i* and *i – 4*.
2	X2	Increment of the *b*-value between events *i – 4* and *i – 8*.
3	X3	Increment of the *b*-value between events *i – 8* and *i – 12*.
4	X4	Increment of the *b*-value between events *i – 12* and *i – 16*.
5	X5	Increment of the *b*-value between events *i – 16* and *i – 20*.
6	X6	Maximum magnitude from the events recorded in the past week (based on Omori–Utsu's law).
7	X7	Probability of recording an event with a magnitude ≥ 6.0, computed using a probability density function.

https://doi.org/10.31436/iiumej.v26i3.3634

All temporal and spatial features are combined into a unified spatiotemporal input matrix. This integrated feature set trains the proposed ConvBiLSTM-Net, a hybrid deep learning model that captures both localized spatial characteristics and long-range temporal dependencies.

The architecture begins with a one-dimensional convolutional layer (256 filters, kernel size = 3), followed by batch normalization and max pooling to extract relevant spatial patterns and suppress noise. This is followed by two stacked bidirectional long short-term memory (BiLSTM) layers (each with 256 units) that model both forward and backward temporal dynamics. Dropout layers (dropout rate = 0.3) are included after each BiLSTM layer to mitigate overfitting. Subsequently, the output is passed through a fully connected dense layer with 128 units, followed by a final dense layer with a single output neuron. This architecture trains separate models for each target variable: magnitude, depth, latitude, and longitude.

Table 3. Additional Temporal Features Extracted from the Earthquake Catalog

#	Feature	Description
1	class_mag	Classification of earthquake magnitude (e.g., small, light, moderate).
2	class_depth	Classification of earthquake depth (e.g., shallow, intermediate, deep).
3	Max magnitude expected	Maximum magnitude expected based on seismic analysis or modeling.
4	Max magnitude observed	Highest magnitude observed in the dataset.
5	Elapsed days	Total number of days elapsed since the first recorded earthquake.
6	Square root of energy	Square root of the energy released by the earthquake, often related to magnitude.
7	Timestamp	Exact time of the earthquake event (date and time).
8	Longitude	Longitude coordinate of the earthquake's epicenter.
9	Latitude	Latitude coordinate of the earthquake's epicenter.
10	Depth	Depth of the earthquake's epicenter
11	Magnitude	Magnitude of the earthquake event.

Hyperparameters are configured: learning rate = 0.001, batch size = 32, and 256 LSTM units. Early stopping based on validation loss is applied to prevent overfitting and ensure generalization. The dataset is split chronologically into 80% for training, 10% for validation, and 10% for testing. Two forecasting strategies are employed to evaluate the model's predictive performance comprehensively.

The short-term prediction task employs a sliding window approach using the most recent 30 earthquake events to forecast the subsequent (one-day-ahead) event. In contrast, the medium-term strategy involves training the model on data from 1974 to 2018 and forecasting earthquake activity for the period 2019–2023. Model performance is evaluated using both regression and classification metrics, including Accuracy, F1 Score, Coefficient of Determination (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE).

2.3. Faults Density Calculation

Anselin extended the Local Moran's I statistic to develop the Local Indicators of Spatial Association (LISA) in 1995 [45]. LISA quantifies the degree of localized spatial clustering of similar values within a dataset, derived using the Local Moran Index. Unlike the global Moran's I, which measures overall spatial autocorrelation, LISA identifies local patterns of spatial dependence [46]. The Local Moran's I statistic is presented in Equation (12).

$$I_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j \neq i}^{n} w_{i,j} \left(x_j - \bar{X}\right) \quad (12)$$

LISA facilitates spatial pattern analysis at a finer scale. If the test result is statistically insignificant, no spatial autocorrelation is present. However, a significant result indicates one

of four spatial patterns: high-high (HH), where high values are clustered together; low-low (LL), where low values cluster; high-low (HL), where low values surround high values; and low-high (LH), where high values surround low values. The sum of the individual LISA values corresponds to the global spatial association measure [47, 48].

Bivariate LISA extends this concept by correlating one variable in a given location with a different variable in neighboring areas. This approach provides a more nuanced spatial analysis, capturing localized dependencies that may not be visible in global measures. As a result, Bivariate LISA enhances the detection of spatial interactions and improves the understanding of regional variations in spatial data. The Bivariate LISA index is given in Equation (13).

$$I_B = \frac{\sum_i \ x_i(\sum_j w_{ij}(d)\, y_j\,)}{\sum_i \left(x_i\ \right)^2} \tag{13}$$

3. RESULTS AND DISCUSSION

3.1. Performance Evaluation of ConvBiLSTM-Net for Multi-Horizon Earthquake Prediction

This section empirically evaluates the proposed ConvBiLSTM-Net model for short-term and medium-term earthquake prediction. The model is tested under two feature configurations: (1) temporal features only, and (2) spatiotemporal features that integrate both spatial and temporal dimensions. The model's performance is assessed using standard evaluation metrics, including accuracy, F1 score, coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE).

Table 4. Performance of the proposed ConvBiLSTM-Net model for short-term earthquake prediction.

Prediction	Feature	Accuracy	F1 Score	R^2	RMSE	MAE
Magnitude	Temporal	0.9075	0.8388	0.7367	0.0392	0.0314
	Spatiotemporal	0.9095	0.8421	0.7835	0.0371	0.0308
Depth	Temporal	0.8489	0.6509	0.8411	0.0579	0.0397
	Spatiotemporal	0.9368	0.9038	0.8595	0.0544	0.0369
Longitude	Temporal	0.7405	0.8204	0.1187	11.8287	8.3620
	Spatiotemporal	0.8773	0.8821	0.7641	5.2055	2.7096
Latitude	Temporal	0.9084	0.8405	0.2297	0.2392	0.1671
	Spatiotemporal	0.9257	0.9607	0.7531	0.1354	0.0763

Table 4 summarizes the performance of the ConvBiLSTM-Net model for short-term earthquake prediction using both temporal and spatiotemporal features. Compared to using temporal features alone, the results demonstrate that incorporating spatial features improves performance across all prediction parameters, magnitude, depth, longitude, and latitude. For magnitude prediction, the spatiotemporal model consistently outperforms the temporal-only model. Accuracy increases by 0.22%, reaching 0.9095, while the F1 score improves by 0.39%, reaching 0.8421. A notable improvement of 6.35% is observed in the coefficient of determination (R^2), indicating a stronger correlation between predicted and actual values. Meanwhile, RMSE and MAE decrease by 5.36% and 1.91%, respectively, reflecting lower prediction errors. These findings suggest incorporating spatial context enhances the model's ability to capture seismic dynamics in short-term scenarios.

https://doi.org/10.31436/iiumej.v26i3.3634

Similarly, the model exhibits substantial performance gains for depth prediction, particularly in the F1 score, which increases by 38.85%. Accuracy also improves by 10.35%, reflecting better classification performance. A moderate improvement of 2.19% is observed in R^2, accompanied by decreases in RMSE and MAE by 6.04% and 7.05%, respectively. These outcomes indicate that, although depth is primarily driven by temporal factors, integrating spatial information enhances the model's generalizability.

Moreover, incorporating spatiotemporal features significantly improves the prediction of earthquake locations, particularly in terms of latitude and longitude. For longitude prediction, the R^2 value increases dramatically by 543.72%, from 0.1187 to 0.7641, demonstrating a substantial enhancement in the model's capacity to explain variability in the data. Accuracy for longitude prediction also improves by 18.47%, reinforcing the model's effectiveness in identifying spatial patterns associated with earthquake epicenters.

Meanwhile, the RMSE decreased by 55.99%, from 11.8287 to 5.2055, and the MAE decreased by 67.61%, from 8.3620 to 2.7096. This reduction in prediction error provides strong evidence that integrating spatial features significantly enhances the model's performance in earthquake longitude prediction. For latitude, the R^2 value increased by 227.86%, from 0.2297 to 0.7531, indicating a substantial improvement in the model's ability to explain data variability. Additionally, the MAE decreased by 54.34%, from 0.1671 to 0.0763, reflecting a considerable reduction in the average prediction error. This was accompanied by a 43.39% decrease in RMSE, from 0.2392 to 0.1354. Furthermore, the accuracy increased by 1.90%, from 0.9084 to 0.9257, reinforcing the reliability of the spatiotemporal model in predicting earthquake latitude. Overall, these results further confirm the effectiveness of the proposed model in capturing the complex spatial dynamics involved in earthquake latitude prediction.

Table 5 presents the earthquake prediction results for the medium-term scenario. These findings reinforce the results presented in Table 4, indicating that the proposed model, which incorporates spatiotemporal features, enhances earthquake prediction performance in the short term. The consistent performance improvement across both temporal scenarios suggests that the ConvBiLSTM-Net model effectively captures seismic dynamics and is reliable for predicting earthquake events over medium-term horizons.

Table 5. Performance of the proposed ConvBiLSTM-Net model for medium-term earthquake prediction

Prediction	Feature	Accuracy	F1 Score	R^2	RMSE	MAE
Magnitude	Temporal	0.9031	0.8231	0.7811	0.1721	0.1381
	Spatiotemporal	0.9458	0.9082	0.9917	0.1721	0.1448
Depth	Temporal	0.9361	0.8876	0.9338	0.3899	0.0271
	Spatiotemporal	0.9388	0.8924	0.9374	0.2418	0.0185
Longitude	Temporal	0.8003	0.8891	0.0878	0.2048	0.2561
	Spatiotemporal	0.9426	0.9632	0.7589	0.2477	0.1661
Latitude	Temporal	0.8623	0.8357	0.6553	0.2584	0.1875
	Spatiotemporal	0.9314	0.9231	0.9201	0.1615	0.1005

Regarding magnitude prediction, the spatiotemporal model shows a notable performance improvement. The F1 score increased by 10.34%, from 0.8231 to 0.9082, while the coefficient of determination (R^2) rose significantly by 26.99%, from 0.7811 to 0.9917, indicating a near-perfect correlation between predicted and actual values. Although the RMSE value remained constant at 0.1721, the MAE slightly increased by 4.85%, from 0.1381 to 0.1448. Moreover, accuracy improved by 4.73%, from 0.9031 to 0.9458, further supporting the model's overall enhancement in classification performance. This overall improvement in model fit,

accompanied by a minor increase in average error, suggests that the increased complexity of medium-term seismic dynamics slightly impacts prediction accuracy.

For depth prediction, the spatiotemporal model also demonstrates meaningful improvements. The R^2 increased by 0.39%, from 0.9338 to 0.9374, while the RMSE and MAE decreased by 37.98% and 31.73%, respectively. Accuracy rose slightly by 0.29%, from 0.9361 to 0.9388, and the F1 score improved by 0.54%, from 0.8876 to 0.8924. The substantial reduction in error metrics and consistent gains in accuracy and F1 score underscore the model's ability to produce more accurate and reliable depth predictions. This improvement suggests that spatial features are crucial in capturing vertical seismic patterns over extended prediction periods.

The most significant improvement is observed in the prediction of longitude coordinates. The R^2 increased dramatically by 764.35%, from 0.0878 to 0.7589. Accuracy rose by 17.78%, from 0.8003 to 0.9426, and the F1 score increased by 8.33%, reaching 0.9632. Although RMSE slightly increased by 20.95%, from 0.2048 to 0.2477, MAE decreased substantially by 35.14%, from 0.2561 to 0.1661. This combination of enhanced accuracy and reduced average error highlights the importance of spatial components in improving the precision of east–west earthquake location predictions.

Similarly, predictions for latitude coordinates exhibit consistent performance enhancements. The R^2 increased by 40.41%, from 0.6553 to 0.9201, reflecting the model's improved capacity to capture the north–south spatial distribution of seismic events. RMSE and MAE declined sharply by 37.50% and 46.40%, respectively, to 0.1615 and 0.1005. Accuracy improved by 8.01%, from 0.8623 to 0.9314, and the F1 score rose by 10.46%, from 0.8357 to 0.9231, further supporting the model's robustness in spatial prediction tasks.

3.2. Comparative Evaluation of ConvBiLSTM-Net and Baseline Architectures for Earthquake Prediction

A comparative evaluation is conducted to benchmark the proposed ConvBiLSTM-Net model against conventional deep learning baselines, including CNN-LSTM, CNN-LSTM with Attention, and CNN-BiLSTM. All models are trained and tested using identical spatiotemporal feature inputs to ensure fairness and consistency. The evaluation focuses on four key seismic prediction targets: earthquake magnitude, depth, and epicentral coordinates (longitude and latitude), across both short-term and medium-term forecasting horizons. Model performance is assessed using a comprehensive set of predictive metrics that capture classification accuracy, regression quality, and overall prediction error.

3.2.1. Model Performance on Short-Term Earthquake Prediction

Table 6 presents the proposed ConvBiLSTM-Net model's magnitude prediction performance compared with three baseline architectures. The results indicate that the proposed model achieves the highest accuracy of 0.9095, representing a relative improvement of approximately 0.25% over the CNN-LSTM, which is the strongest baseline in terms of accuracy. Additionally, the F1 score reaches 0.8421, marking a 0.53% improvement over the average F1 score of the baseline models, and demonstrating the model's enhanced ability to balance precision and recall.

Notable improvements are also observed in the regression metrics. The proposed model achieves the lowest RMSE (0.0371) and MAE (0.0308), corresponding to a reduction in prediction error of approximately 24.45% and 9.14%, respectively, compared to the worst-performing baseline (CNN-LSTM). When compared to the best-performing baseline in terms

https://doi.org/10.31436/iiumej.v26i3.3634

of regression (CNN-BiLSTM), ConvBiLSTM-Net still demonstrates a significant reduction in error. Furthermore, the R^2 score of ConvBiLSTM-Net increases substantially to 0.7835, compared to the baseline range of 0.6231 to 0.6771. This improvement, ranging from 15.71% to 25.74%, reflects a stronger correlation between the predicted and actual magnitude values.

These results confirm the proposed architecture's superiority in capturing complex nonlinear relationships and spatiotemporal patterns in seismic data.

Table 6. Magnitude Prediction

Model	Accuracy	F1 Score	RMSE	MAE	R^2
CNN-LSTM	0.9072	0.8378	0.0491	0.0339	0.6231
CNN-LSTM with Attention	0.9071	0.8376	0.0466	0.0325	0.6578
CNN-BiLSTM	0.9071	0.8376	0.0453	0.0317	0.6771
ConvBiLSTM-Net (Proposed Model)	0.9095	0.8421	0.0371	0.0308	0.7835

Table 7 presents the performance of the proposed ConvBiLSTM-Net model in predicting earthquake depth compared to three baseline architectures. Regarding regression metrics, the proposed model achieved the lowest RMSE of 0.0544 and MAE of 0.0369, indicating a consistent improvement in prediction accuracy. Specifically, the RMSE decreased by 5.06% compared to CNN-LSTM, 3.03% compared to CNN-LSTM with Attention, and 1.81% compared to CNN-BiLSTM. Similarly, the MAE showed a significant reduction, ranging from 3.40% to 7.05% compared to the baseline models. These reductions in error metrics highlight the ConvBiLSTM-Net's advantage in minimizing deviations from actual values.

Further improvements are evident in the classification metrics. The ConvBiLSTM-Net model achieved an accuracy of 0.9368 and an F1 score of 0.9038, corresponding to relative improvements of 7.70% and 14.97%, respectively, compared to the average performance of the baseline models. Additionally, the R^2 value increased to 0.8595, representing a 1.13% improvement and indicating a stronger correlation between predicted and actual depth values.

Overall, these results reinforce the effectiveness of the ConvBiLSTM-Net model in accurately and reliably capturing spatiotemporal patterns essential for hypocenter depth estimation, outperforming conventional deep learning architectures.

Table 7. Depth Prediction

Model	Accuracy	F1 Score	RMSE	MAE	R^2
CNN-LSTM	0.8582	0.8442	0.0573	0.0382	0.8442
CNN-LSTM With Attention	0.8691	0.8511	0.0561	0.0381	0.8511
CNN-BiLSTM	0.8821	0.6631	0.0554	0.0397	0.8545
ConvBiLSTM-Net (Proposed Model)	0.9368	0.9038	0.0544	0.0369	0.8595

In line with the observed improvements in depth prediction, the ConvBiLSTM-Net model also demonstrates superior performance in predicting earthquake longitude, as shown in Table 8 The proposed model achieves the highest accuracy of 0.8773 and an F1 score of 0.8821, outperforming the average of the three baseline models by 11.61% and 17.17%, respectively. These results indicate a notable enhancement in the model's classification capability for predicting earthquake longitude.

Regarding regression metrics, ConvBiLSTM-Net also achieves the highest R^2 value of 0.7641, reflecting a stronger ability to capture variations in the longitudinal patterns of

earthquake locations. This increase in R^2 represents improvements of 16.34%, 11.94%, and 4.37% over CNN-BiLSTM, CNN-LSTM, and CNN-LSTM with Attention, respectively. While the RMSE (0.1939) and MAE (0.1327) achieved by the ConvBiLSTM-Net are comparable to those of the CNN-LSTM with Attention model, both values are higher than those produced by CNN-BiLSTM. This suggests that, although ConvBiLSTM-Net excels in correlation-based performance, it has not yet fully outperformed the baseline in minimizing absolute error for longitude estimation.

Nevertheless, integrating bidirectional temporal modeling with spatial feature extraction substantially enhances the model‘s overall predictive capability, reinforcing its effectiveness in earthquake location estimation.

Table 8. Longitude Prediction

Model	Accuracy	F1 Score	RMSE	MAE	R^2
CNN-LSTM	0.7752	0.7396	0.1684	0.1069	0.6826
CNN-LSTM With Attention	0.8258	0.8106	0.1939	0.1327	0.7321
CNN-BiLSTM	0.7572	0.7084	0.1562	0.0961	0.6568
ConvBiLSTM-Net (Proposed Model)	0.8773	0.8821	0.1939	0.1327	0.7641

Table 9 confirms that the proposed ConvBiLSTM-Net model outperforms all baseline models across the evaluated metrics for latitude prediction. The model achieves the highest accuracy of 0.9257, representing improvements of 0.64% over CNN-LSTM, 1.18% over CNN-LSTM with Attention, and 4.93% over CNN-BiLSTM. Furthermore, the F1 score reaches 0.9607, reflecting relative improvements ranging from 0.26% to 2.57% compared to the baseline models. The ConvBiLSTM-Net also yields the lowest RMSE of 0.1354, corresponding to a reduction of 14.90% from CNN-LSTM, 3.77% from CNN-LSTM with Attention, and 4.18% from CNN-BiLSTM. Similarly, the MAE decreases significantly to 0.0763, showing reductions of 35.94%, 24.60%, and 20.27% compared to CNN-LSTM, CNN-LSTM with Attention, and CNN-BiLSTM, respectively. The R^2 value increases to 0.7531, representing an improvement ranging from approximately 2.69% to 14.16%, further demonstrating the superiority of the ConvBiLSTM-Net in accurately modeling the spatial distribution of earthquake latitude coordinates.

Table 9. Latitude Prediction

Model	Accuracy	F1 Score	RMSE	MAE	R^2
CNN-LSTM	0.9198	0.9582	0.1591	0.1191	0.6597
CNN-LSTM With Attention	0.9149	0.9554	0.1407	0.1012	0.7334
CNN-BiLSTM	0.8822	0.9366	0.1413	0.0957	0.7313
ConvBiLSTM-Net (Proposed Model)	0.9257	0.9607	0.1354	0.0763	0.7531

3.2.2. Model Performance on Medium-Term Earthquake Prediction

Table 10 presents the proposed ConvBiLSTM-Net model's performance in predicting earthquake magnitude under medium-term forecasting scenarios. The model achieves an accuracy of 0.9458, reflecting an improvement of approximately 4.73% over the best-performing baseline. Likewise, the F1-score increases to 0.9082, showing a relative

https://doi.org/10.31436/iiumej.v26i3.3634

improvement of 10.30% to 10.51% compared to the baseline values (ranging from 0.8217 to 0.8231).

The model records an RMSE of 0.1721 for regression metrics, representing a prediction error reduction of approximately 7.08% to 8.87% compared to the baselines. The coefficient of determination (R^2) improves significantly to 0.8317, corresponding to an increase of 8.21% to 13.27% relative to baseline models. This elevated R^2 indicates a stronger correlation between predicted and actual magnitudes, suggesting enhanced model capability in capturing data variability.

Although the MAE value of 0.1448 is slightly higher than the most competitive baseline (0.1334), the relative increase, ranging from 1.83% to 8.57%, is marginal. This slight increase in MAE is outweighed by the substantial improvements across other key performance indicators, particularly in classification and regression accuracy. Overall, the ConvBiLSTM-Net demonstrates a robust and balanced architecture for medium-term magnitude forecasting by effectively leveraging spatiotemporal dependencies to enhance predictive accuracy and generalization.

Table 10. Magnitude Prediction

Model	Accuracy	F1 Score	RMSE	MAE	R^2
CNN-LSTM	0.9023	0.8218	0.1858	0.1391	0.7425
CNN-LSTM With Attention	0.9025	0.8217	0.1888	0.1422	0.7342
CNN-BiLSTM	0.9031	0.8231	0.1862	0.1334	0.7686
ConvBiLSTM-Net (Proposed Model)	0.9458	0.9082	0.1721	0.1448	0.8317

Table 11 presents the performance of the ConvBiLSTM-Net model in predicting earthquake focal depth under medium-term forecasting scenarios. The model achieves an accuracy of 0.9388, reflecting an improvement of approximately 2.24% over the weakest baseline (CNN-LSTM with Attention). Additionally, the F1-score improves to 0.8924, indicating a relative increase of 2.89% to 3.27% over the baseline models. This improvement highlights the model's enhanced reliability in classifying depth categories.

Regarding regression, the model achieves a mean absolute error (MAE) of 0.0185, representing a reduction of 17.04% to 29.12% compared to the baselines. This substantial decrease in MAE indicates the model's superior ability to minimize absolute prediction errors in seismic depth estimation. Furthermore, the coefficient of determination (R^2) increases to 0.9374, corresponding to an improvement of 4.91% to 9.58%. This reflects the model's enhanced capacity to capture variance in depth values and ensure a closer fit between predicted and actual outcomes.

Table 11. Depth Prediction

Model	Accuracy	F1 Score	RMSE	MAE	R^2
CNN-LSTM	0.9188	0.8653	0.3675	0.0261	0.8555
CNN-LSTM With Attention	0.9182	0.8641	0.3526	0.0244	0.8671
CNN-BiLSTM	0.9237	0.8673	0.0481	0.0223	0.8935
ConvBiLSTM-Net (Proposed Model)	0.9388	0.8924	0.2418	0.0185	0.9374

While the RMSE value of 0.2418 is considerably higher than that of the best-performing baseline model, this outcome can be attributed to the inherent sensitivity of the RMSE metric

https://doi.org/10.31436/iiumej.v26i3.3634

to outliers, particularly infrequent deep-focus seismic events. Nevertheless, the consistent improvements in classification performance and the substantial reduction in MAE suggest that the proposed ConvBiLSTM-Net model yields more reliable and practically meaningful depth estimations. Such reliability is especially critical in operational scenarios where minimizing average predictive error is more important than accounting for isolated deviations. Together, these results reinforce the robustness and generalization capability of the proposed architecture in medium-term seismic depth forecasting.

Table 12 presents the performance of the proposed ConvBiLSTM-Net model in predicting earthquake longitude under medium-term scenarios. The ConvBiLSTM-Net achieves an accuracy of 0.9426, representing an improvement ranging from approximately 1.49% to 4.18% compared to the baseline models, with the highest baseline accuracy observed in the CNN-BiLSTM (0.9288). The F1 score also increases to 0.9632, reflecting a relative improvement of approximately 0.99% to 2.67%. This enhancement demonstrates the model's stronger classification performance in distinguishing spatial categories along the longitudinal axis. Furthermore, the coefficient of determination (R^2) improves to 0.7589, representing a relative gain of approximately 6.27% to 17.56% over the baseline models. The higher R^2 value indicates the model's enhanced ability to capture spatial variance in earthquake location data. These improvements underscore the ConvBiLSTM-Net's increased capability to model and interpret spatial dependencies that characterize earthquake longitude.

Table 12. Longitude Prediction

Model	Accuracy	F1 Score	RMSE	MAE	R^2
CNN-LSTM	0.9259	0.9521	0.1684	0.1069	0.6642
CNN-LSTM With Atten	0.9048	0.9382	0.2166	0.1607	0.6454
CNN-BiLSTM	0.9288	0.9539	0.1562	0.0961	0.7142
ConvBiLSTM-Net (Proposed Model)	0.9426	0.9632	0.2477	0.1661	0.7589

However, higher absolute prediction errors accompany these gains in classification accuracy and variance explanation. The RMSE increases to 0.2477, and the MAE rises to 0.1661, corresponding to relative increases of approximately 58.58% and 72.84%, respectively, when compared to the best-performing baseline (CNN-BiLSTM), which recorded an RMSE of 0.1562 and an MAE of 0.0961. These results suggest that while the model excels at identifying spatial patterns, it may be more susceptible to larger individual prediction deviations. The elevated error values could stem from the model's sensitivity to positional outliers or broader spatial variability in longitudinal distribution. Despite this trade-off, the ConvBiLSTM-Net demonstrates strong predictive capabilities in capturing longitudinal patterns, particularly regarding accuracy, F1 score, and explained variance. While further refinement may be necessary to reduce point-specific errors, the model remains robust and well-suited for medium-term seismic spatial forecasting, especially in applications where pattern recognition and classification fidelity are prioritized.

Table 13 summarizes the medium-term latitude prediction results, indicating that the proposed ConvBiLSTM-Net establishes a new benchmark for overall performance. The model achieves an accuracy of 0.9314, representing an improvement of approximately 2.59% over the best-performing baseline model, CNN-BiLSTM (0.9078). Regarding classification quality, the F1 score rises to 0.8942, reflecting a 5.31% enhancement over the strongest baseline score of 0.8491, recorded by CNN-LSTM with Attention. The proposed model also shows a modest gain in explanatory power, with the coefficient of determination (R^2) increasing to 0.9201, equivalent to a 1.30% improvement over the CNN-BiLSTM baseline (R^2 = 0.9083). This

https://doi.org/10.31436/iiumej.v26i3.3634

increase suggests a more substantial alignment between the predicted and observed spatiotemporal patterns.

A slight trade-off in absolute prediction errors accompanies these improvements. The RMSE increases to 0.1615, approximately 12.80% higher than the lowest baseline error of 0.1432 recorded by CNN-BiLSTM. Similarly, the MAE rises marginally to 0.1005, representing an increase of only 0.70% compared to the lowest baseline MAE, also achieved by CNN-BiLSTM (0.0998). Despite these modest increases in absolute error, the substantial gains in accuracy, F1 score, and R^2 underscore the ConvBiLSTM-Net model's superior ability to capture complex seismic patterns along the latitude axis. Overall, the proposed model offers the most reliable balance between classification fidelity and regression accuracy, positioning it as the most effective architecture for medium-term latitude forecasting in this study.

Table 13. Latitude Prediction

Model	Accuracy	F1 Score	RMSE	MAE	R^2
CNN-LSTM	0.8961	0.8263	0.1595	0.1209	0.9009
CNN-LSTM With Atten	0.8989	0.8491	0.2081	0.2371	0.8488
CNN-BiLSTM	0.9078	0.8172	0.1432	0.0998	0.9083
ConvBiLSTM-Net (Proposed Model)	0.9314	0.8942	0.1615	0.1005	0.9201

4. CONCLUSION

This study demonstrates that the proposed ConvBiLSTM-Net model significantly outperforms various deep learning architectures (baselines) in earthquake prediction across all key targets, namely, magnitude, depth, and epicentral coordinates (longitude and latitude), in both short-term and medium-term scenarios. In the short-term prediction scenario, ConvBiLSTM-Net consistently surpasses all baseline models. Specifically, for magnitude prediction, the model achieves a 20.05% improvement in R^2, alongside increases of 0.26% and 0.53% in accuracy and F1 score, respectively. Additionally, it reduces RMSE and MAE by 21.06% and 5.81%, reflecting enhanced precision in magnitude estimation. For depth prediction, ConvBiLSTM-Net improves accuracy by 7.70% and F1 score by 14.97%, with a modest R^2 gain of 1.13%. It also lowers RMSE and MAE by 3.32% and 4.57%, respectively, indicating overall improvement in depth prediction.

Beyond magnitude and depth, ConvBiLSTM-Net also performs strongly in epicentral coordinate prediction. For latitude prediction, the model increases R^2 by 6.35% and accuracy by 2.22%, while substantially decreasing RMSE and MAE by 7.91% and 27.56%, respectively, indicating significant spatial improvement along the north–south axis. For longitude prediction, it achieves an 11.61% gain in accuracy and a 10.66% increase in R^2. However, RMSE and MAE increase by 12.19% and 18.59%, respectively, suggesting that while the model better captures variance in longitude (as reflected by the higher R^2), it also yields higher absolute prediction errors than the baselines.

ConvBiLSTM-Net maintains superior performance across all key targets in the medium-term prediction scenario. The model achieves an R^2 improvement of 11.13% for magnitude prediction, with increased accuracy and F1 score of 4.78% and 10.46%, respectively. RMSE decreases by 7.94%, although MAE increases slightly by 4.75%, indicating better variance explanation with a minor trade-off in localized error. Regarding depth prediction, the model improves accuracy by 2.02% and F1 score by 3.10%, with a notable R^2 gain of 7.50%. It also

significantly reduces RMSE and MAE by 5.58% and 23.76%, respectively, further underscoring its ability to estimate depth precisely.

Turning to spatial prediction, the model also demonstrates notable improvements in epicentral coordinate estimation. For latitude prediction, it improves R^2 by 3.85% and accuracy by 3.38%, while significantly reducing the mean absolute error (MAE) by 34.14% and the root mean square error (RMSE) by 5.15%. These results indicate meaningful enhancements in spatial prediction accuracy along the north–south axis. For longitude prediction, ConvBiLSTM-Net achieves an R^2 improvement of 12.50%, with a 2.48% increase in accuracy and a 1.60% gain in F1 score. However, these improvements come at the cost of increased error, with RMSE and MAE rising by 37.31% and 37.01%, respectively. This suggests that although the model more effectively captures longitudinal variance, it also yields higher absolute prediction errors.

Based on these findings, ConvBiLSTM-Net demonstrates strong capabilities in modeling spatiotemporal relationships while maintaining robust and superior prediction performance across different time horizons. The model consistently enhances the accuracy of predicting all key seismic parameters, magnitude, depth, and particularly the earthquake epicenter (latitude and longitude). Notably, while absolute errors in longitude prediction increase in some instances, the model still exhibits improved pattern recognition, as indicated by higher R^2 values. By leveraging the synergy between spatial and temporal features, ConvBiLSTM-Net effectively captures complex seismic dynamics, resulting in improved generalization in both short-term and medium-term scenarios. These findings confirm its potential as a practical and adaptive solution for advancing earthquake prediction systems, particularly in enhancing epicenter localization, which is critical for early warning and disaster risk mitigation efforts.

ACKNOWLEDGEMENT

This work was supported by the Indonesian Education Scholarship (*Beasiswa Pendidikan Indonesia*, BPI), administered by the Center for Higher Education Funding and Assessment, Ministry of Higher Education, Science, and Technology of the Republic of Indonesia (PPAPT), and the Indonesia Endowment Fund for Education (*Lembaga Pengelola Dana Pendidikan*, LPDP), as acknowledged in Decree No. 02462/BPPT/BPI.06/9/2024. The authors also thank the Laboratory of Computer Systems and Networks, Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, for providing technical facilities and academic support during this research.

REFERENCES

[1] Alarifi, A.S.N., Alarifi, N.S.N., Al-Humidan, S.: Earthquakes magnitude predication using artificial neural network in northern Red Sea area. J King Saud Univ Sci. 24, 301–313 (2012). https://doi.org/10.1016/j.jksus.2011.05.002.

[2] Berhich, A., Belouadha, F.-Z., Kabbaj, M.I.: An attention-based LSTM network for large earthquake prediction. Soil Dynamics and Earthquake Engineering. 165, 107663 (2023). https://doi.org/10.1016/j.soildyn.2022.107663.

[3] Shidik, G.F., Pramunendar, R.A., Kusuma, E.J., Saraswati, G.W., Winarsih, N.A.S., Rohman, M.S., Saputra, F.O., Naufal, M., Andono, P.N.: LUTanh Activation Function to Optimize BI-LSTM in Earthquake Forecasting. International Journal of Intelligent Engineering and Systems. 17, 572 – 583 (2024). https://doi.org/10.22266/ijies2024.0229.48.

[4] Maharani, E.: Infografis Kerusakan Ekonomi Akibat Gempa Besar di Dunia, https://visual.republika.co.id/berita/rrsyo1335/infografis-kerusakan-ekonomi-akibat-gempa-besar-di-dunia, last accessed 2023/10/01.

[5] UYEDA, S.: On Earthquake Prediction in Japan. Proceedings of the Japan Academy, Series B. 89, 391–400 (2013). https://doi.org/10.2183/pjab.89.391.

[6] HAYAKAWA, M., SCHEKOTOV, A., POTIRAKIS, S., EFTAXIAS, K.: Criticality features in ULF magnetic fields prior to the 2011 Tōhoku earthquake. Proceedings of the Japan Academy, Series B. 91, 25–30 (2015). https://doi.org/10.2183/pjab.91.25.

[7] Alimoradi, H., Rahimi, H., Ovaisi Moaakhar, M.: Investigation of changes in the cumulative number of magnetic anomalies before and after earthquakes using satellite data. Advances in Space Research. 75, 895–907 (2025). https://doi.org/10.1016/j.asr.2024.09.012.

[8] Wang, Q., Guo, Y., Yu, L., Li, P.: Earthquake Prediction Based on Spatio-Temporal Data Mining: An LSTM Network Approach. IEEE Trans Emerg Top Comput. 8, 148–158 (2017). https://doi.org/10.1109/TETC.2017.2699169.

[9] Kagan, Y.Y., Jackson, D.D.: Long-Term Earthquake Clustering. Geophys J Int. 104, 117–134 (1991). https://doi.org/10.1111/j.1365-246X.1991.tb02498.x.

[10] Kagan, Y.Y., Jackson, D.D.: Long-term probabilistic forecasting of earthquakes. J Geophys Res Solid Earth. 99, 13685–13700 (1994). https://doi.org/10.1029/94JB00500.

[11] Banna, Md.H. Al, Taher, K.A., Kaiser, M.S., Mahmud, M., Rahman, Md.S., Hosen, A.S.M.S., Cho, G.H.: Application of Artificial Intelligence in Predicting Earthquakes: State-of-the-Art and Future Challenges. IEEE Access. 8, 192880–192923 (2020). https://doi.org/10.1109/ACCESS.2020.3029859.

[12] Mojtahedi, F.F., Yousefpour, N., Chow, S.H., Cassidy, M.: Deep Learning for Time Series Forecasting: Review and Applications in Geotechnics and Geosciences. Archives of Computational Methods in Engineering. (2025). https://doi.org/10.1007/s11831-025-10244-5.

[13] Majumder, R.Q.: Machine Learning for Predictive Analytics: Trends and Future Directions. SSRN Electronic Journal. (2025). https://doi.org/10.2139/ssrn.5267273.

[14] Jamarani, A., Haddadi, S., Sarvizadeh, R., Haghi Kashani, M., Akbari, M., Moradi, S.: Big data and predictive analytics: A systematic review of applications. Artif Intell Rev. 57, 176 (2024). https://doi.org/10.1007/s10462-024-10811-5.

[15] Huang, C.-J., Kuo, P.-H.: A Short-Term Wind Speed Forecasting Model by Using Artificial Neural Networks with Stochastic Optimization for Renewable Energy Systems. Energies (Basel). 11, 2777 (2018). https://doi.org/10.3390/en11102777.

[16] Peng, L., Wang, L., Ai, X.-Y., Zeng, Y.-R.: Forecasting Tourist Arrivals via Random Forest and Long Short-term Memory. Cognit Comput. 13, 125–138 (2021). https://doi.org/10.1007/s12559-020-09747-z.

[17] Li, X., Wu, P.: Stock Price Prediction Incorporating Market Style Clustering. Cognit Comput. 14, 149–166 (2022). https://doi.org/10.1007/s12559-021-09820-1.

[18] Wang, X., Nguyen, K., Nguyen, B.P.: Churn Prediction using Ensemble Learning. In: Proceedings of the 4th International Conference on Machine Learning and Soft Computing. pp. 56–60. ACM, New York, NY, USA (2020). https://doi.org/10.1145/3380688.3380710.

[19] Asim, K.M., Martínez-Álvarez, F., Basit, A., Iqbal, T.: Earthquake magnitude prediction in Hindukush region using machine learning techniques. Natural Hazards. 85, 471–486 (2017). https://doi.org/10.1007/s11069-016-2579-3.

[20] Asim, K.M., Idris, A., Iqbal, T., Martínez-Álvarez, F.: Earthquake prediction model using support vector regressor and hybrid neural networks. PLoS One. 13, e0199004 (2018). https://doi.org/10.1371/journal.pone.0199004.

[21] Yaghmaei-Sabegh, S., Tsang, H.-H.: A new site classification approach based on neural networks. Soil Dynamics and Earthquake Engineering. 31, 974–981 (2011). https://doi.org/10.1016/j.soildyn.2011.03.004.

[22] Asim, K.M., Moustafa, S.SR., Niaz, I.A., Elawadi, E.A., Iqbal, T., Martínez-Álvarez, F.: Seismicity analysis and machine learning models for short-term low magnitude seismic activity

predictions in Cyprus. Soil Dynamics and Earthquake Engineering. 130, 105932 (2020). https://doi.org/10.1016/j.soildyn.2019.105932.

[23] Jain, R., Nayyar, A., Arora, S., Gupta, A.: A comprehensive analysis and prediction of earthquake magnitude based on position and depth parameters using machine and deep learning models. Multimed Tools Appl. 80, 28419–28438 (2021). https://doi.org/10.1007/s11042-021-11001-z.

[24] Gonzalez, J., Yu, W., Telesca, L.: Gated Recurrent Units Based Recurrent Neural Network for Forecasting the Characteristics of the Next Earthquake. Cybern Syst. 53, 209–222 (2022). https://doi.org/10.1080/01969722.2021.1981637.

[25] Kavianpour, P., Kavianpour, M., Ramezani, A.: Deep Multi-scale Dilated Convolution Neural Network with Attention Mechanism: A Novel Method for Earthquake Magnitude Classification. In: 2022 8th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS). pp. 1–6 (2022). https://doi.org/10.1109/ICSPIS56952.2022.10043978.

[26] Cai, Y., Shyu, M.-L., Tu, Y.-X., Teng, Y.-T., Hu, X.-X.: Anomaly detection of earthquake precursor data using long short-term memory networks. Applied Geophysics. 16, 257–266 (2019). https://doi.org/10.1007/s11770-019-0774-1.

[27] Fabregas, A.C., Arellano, P.B. V., Pinili, A.N.D.: Long-Short Term Memory (LSTM) Networks with Time Series and Spatio-Temporal Approaches Applied in Forecasting Earthquakes in the Philippines. In: Proceedings of the 4th International Conference on Natural Language Processing and Information Retrieval. pp. 188–193. ACM, New York, NY, USA (2020). https://doi.org/10.1145/3443279.3443288.

[28] Berhich, A., Belouadha, F.-Z., Kabbaj, M.I.: An attention-based LSTM network for large earthquake prediction. Soil Dynamics and Earthquake Engineering. 165, 107663 (2023). https://doi.org/10.1016/j.soildyn.2022.107663.

[29] Kail, R., Burnaev, E., Zaytsev, A.: Recurrent Convolutional Neural Networks Help to Predict Location of Earthquakes. IEEE Geoscience and Remote Sensing Letters. 19, 1–5 (2022). https://doi.org/10.1109/LGRS.2021.3107998.

[30] Bhandarkar, T., K, V., Satish, N., Sridhar, S., Sivakumar, R., Ghosh, S.: Earthquake trend prediction using long short-term memory RNN. International Journal of Electrical and Computer Engineering (IJECE). 9, 1304 (2019). https://doi.org/10.11591/ijece.v9i2.pp1304-1312.

[31] Kavianpour, P., Kavianpour, M., Jahani, E., Ramezani, A.: A CNN-BiLSTM model with attention mechanism for earthquake prediction. J Supercomput. 80, 2913–2913 (2024). https://doi.org/10.1007/s11227-023-05497-5.

[32] Kavianpour, P., Kavianpour, M., Jahani, E., Ramezani, A.: Earthquake Magnitude Prediction using Spatia-Temporal Features Learning Based on Hybrid CNN-BiLSTM Model. In: Proceedings - 2021 7th International Conference on Signal Processing and Intelligent Systems, ICSPIS 2021. Institute of Electrical and Electronics Engineers Inc. (2021). https://doi.org/10.1109/ICSPIS54653.2021.9729358.

[33] Geng, Y., Su, L., Jia, Y., Han, C.: Seismic Events Prediction Using Deep Temporal Convolution Networks. Journal of Electrical and Computer Engineering. 2019, 1–14 (2019). https://doi.org/10.1155/2019/7343784.

[34] Yousefzadeh, M., Hosseini, S.A., Farnaghi, M.: Spatiotemporally explicit earthquake prediction using deep neural network. Soil Dynamics and Earthquake Engineering. 144, 106663 (2021). https://doi.org/10.1016/j.soildyn.2021.106663.

[35] Hetényi, G., Epard, J.-L., Colavitti, L., Hirzel, A.H., Kiss, D., Petri, B., Scarponi, M., Schmalholz, S.M., Subedi, S.: Spatial relation of surface faults and crustal seismicity: a first comparison in the region of Switzerland. Acta Geodaetica et Geophysica. 53, 439–461 (2018). https://doi.org/10.1007/s40328-018-0229-9.

[36] Matsuda, T.: Estimation of Future Destructive Earthquakes from Active Faults on Land in Japan. Journal of Physics of the Earth. 25, S251–S260 (1977). https://doi.org/10.4294/jpe1952.25.Supplement_S251.

[37] Matsuda, T.: Active Faults and Damaging Earthquakes in Japan-Macroseismic Zoning and Precaution Fault Zones. (2013). https://doi.org/10.1029/ME004p0279.

[38] Sharma, A., Ahuja, A., Devi, S., Pasari, S.: Use of Spatio-temporal Features for Earthquake Forecasting of imbalanced Data. In: 2022 International Conference on Intelligent Innovations in Engineering and Technology (ICIIET). pp. 178–182. IEEE (2022). https://doi.org/10.1109/ICIIET55458.2022.9967687.

[39] Song, W., Wang, C., Chen, W., Zhang, X., Li, H., Li, J.: Unlocking the spatial heterogeneous relationship between Per Capita GDP and nearby air quality using bivariate local indicator of spatial association. Resour Conserv Recycl. 160, 104880 (2020). https://doi.org/10.1016/j.resconrec.2020.104880.

[40] Phan, H., Andreotti, F., Cooray, N., Chen, O.Y., De Vos, M.: DNN Filter Bank Improves 1-Max Pooling CNN for Single-Channel EEG Automatic Sleep Stage Classification. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 453–456. IEEE (2018). https://doi.org/10.1109/EMBC.2018.8512286.

[41] Qiao, M., Yan, S., Tang, X., Xu, C.: Deep Convolutional and LSTM Recurrent Neural Networks for Rolling Bearing Fault Diagnosis Under Strong Noises and Variable Loads. IEEE Access. 8, 66257–66269 (2020). https://doi.org/10.1109/ACCESS.2020.2985617.

[42] PuSGeN: Peta Sumber Daya dan Bahaya Gempa Indonesia Tahun 2017. Puslitbang PUPR, Indonesia (2017).

[43] Adeli, H., Panakkat, A.: A probabilistic neural network for earthquake magnitude prediction. Neural Networks. 22, 1018–1024 (2009). https://doi.org/10.1016/j.neunet.2009.05.003.

[44] Reyes, J., Morales-Esteban, A., Martínez-Álvarez, F.: Neural networks to predict earthquakes in Chile. Appl Soft Comput. 13, 1314–1328 (2013). https://doi.org/10.1016/j.asoc.2012.10.014.

[45] Anselin, L.: Local Indicators of Spatial Association, LISA. Geogr Anal. 27, 93–115 (1995). https://doi.org/10.1111/j.1538-4632.1995.tb00338.x.

[46] Lee, J., Wong, D.W.S.: Statistical Analysis with ArcView GIS. John Wiley & Sons, New York (2001).

[47] Mathur, M.: Spatial autocorrelation analysis in plant population: An overview. Journal of Applied and Natural Science. 7, 501–513 (2015). https://doi.org/10.31018/jans.v7i1.639.

[48] Gao, Y., Cheng, J., Meng, H., Liu, Y.: Measuring spatio-temporal autocorrelation in time series data of collective human mobility. Geo-spatial Information Science. 22, 166–173 (2019). https://doi.org/10.1080/10095020.2019.1643609.

https://doi.org/10.31436/iiumej.v26i3.3646

MINIATURIZED L-SHAPED AND U-SHAPED RESONATOR BASED 8-BIT AND 12-BIT CHIPLESS RFID TAG

SANAA SALAMA[1*], OMAR TAMIMI[2], MUNERRA ALTAYEB[3], AHLAM ABUZAHEW[4]

[1]*Department of Biomedical Engineering, Arab American University, Jenin, Palestine*
[2]*Department of Electrical Engineering, Najah National University, Nablus, Palestine*
[3]*Dept. of Communications and Computer Eng, Al-Ahliyya Amman University, Amman, Jordan*
[4]*Dept. of Electrical & Electronics Eng, Sakarya University of Applied Sciences, Sakarya, Turkey*

**Corresponding author: sanaa.salama@aaup.edu*

(Received: 9 March 2025; Accepted: 29 July 2025; Published online: 9 September 2025)

ABSTRACT: Chipless Radio Frequency Identification (RFID) technology is a wireless technology that uses radio frequency signals (RF) to identify objects automatically. Chipless RFID tags promise a low-cost and printable solution for item tracking, authentication, and sensing in various applications, including supply chain management and the Internet of Things (IoT). In this work, two compact-sized prototypes of 8-bit chipless radio frequency identification (RFID) tags are modeled as L-shaped and U-shaped resonators. The proposed tags are printed on (14.5mm × 14.5mm) Rogers RT5880 substrate with a dielectric constant, $\varepsilon_r = 2.2$, and a thickness, $h = 1.575\ mm$. The simulated RCS responses for both the L-shaped and the U-shaped 8-bit chipless RFID tags are presented. A prototype of 12 back-to-back L-shaped resonators with different lengths corresponding to resonant frequencies between 5 and 8.5 GHz is proposed to increase the chipless RFID tag capacity. The proposed 12 back-to-back L-shaped resonators chipless RFID tag is printed on (15mm × 25mm) Rogers RT5880 substrate. The simulated RCS response for the compact 12-bit L-shaped resonator-based chipless RFID tag is calculated.

ABSTRAK: Teknologi Pengenalan Frekuensi Radio Tanpa Cip (Chipless RFID) merupakan teknologi tanpa wayar yang menggunakan isyarat frekuensi radio (RF) bagi mengenal pasti objek secara automatik. Tag RFID tanpa wayar menawarkan penyelesaian kos rendah yang berpotensi bagi menjejak item, pengesahan, dan pengesanan dalam pelbagai aplikasi termasuk pengurusan rantaian bekalan dan Internet Benda (IoT). Kajian ini membentangkan dua prototaip bersaiz kompak; iaitu tag RFID 8-bit tanpa wayar yang dimodelkan sebagai resonator berbentuk-L dan berbentuk-U. Tag yang dicadangkan ini dicetak pada substrat Rogers RT5880 (14.5 mm × 14.5 mm) dengan pemalar dielektrik, , dan ketebalan, h = 1.575 mm. Respons simulasi RCS bagi kedua-dua tag 8-bit berbentuk-L dan berbentuk-U dibentangkan dalam kajian ini. Bagi meningkatkan kapasiti tag RFID tanpa wayar, satu prototaip yang terdiri daripada 12 resonator berturutan berbentuk-L dengan panjang berbeza dan berfrekuensi resonan antara 5 hingga 8.5 GHz telah dicadangkan. Tag RFID tanpa wayar ini dicetak pada substrat Rogers RT5880 (15 mm × 25 mm). Respons simulasi RCS bagi tag kompak RFID 12-bit tanpa wayar turut dikira.

KEYWORDS: *Chipless RFID, L-shaped resonator, U-shaped resonator, compact size, Radar cross section (RCS), reading distance.*

1. INTRODUCTION

Radio frequency identification (RFID) is a wireless technology that automatically uses radio frequency signals (RF) to automatically identify objects. An RFID system comprises

https://doi.org/10.31436/iiumej.v26i3.3646

three main components: an RFID tag in which the data is encoded, an RFID reader used to encode the data extracted from the tags, and a host computer, Figure 1, [1]. RFID is commonly used in numerous applications such as security, asset tracking, automatic product identification, and library tracking books. RFID is gradually replacing the traditional barcode. However, the challenge in appropriately using RFID technology instead of conventional barcodes is the cost of the tag [2]. The tag cost depends on the cost of its IC. Thus, efforts have been concentrated on developing chipless RFID tags for low-cost communication systems [3], compact size, and high capacity. In [1], chipless RFID tags are classified into three categories: spectrum-signature-based, amplitude-phase-backscatter-modulation-based, and time domain reflectometry-based chipless RFID tags. Data is encoded into the spectrum using resonant structures in the spectrum-signature-based RFID tags. These tags have several advantages, including being fully printable, having a greater capacity rate, and being inexpensive.

Two techniques are used to design chipless RFID tags in the frequency domain: planar circuits [4] and LC resonators [5-8]. For instance, in [4], an array of identical 18 mm planar dipoles printed on TLY-5 substrate, supplied by Taconics with a thickness of 0.53 mm, a dielectric constant of 2.2, and (tan delta = 0.001), was capacitively tuned to operate between 5 and 6 GHz. Microstrip TL resonator structures were later introduced to increase bit capacity and obtain a compact size. A 6-bit 30 mm x 35 mm meandered microstrip TL tag operating from 3 to 5 GHz was presented in [5] as a chipless RFID tag printed on Taconic substrate with a dielectric constant of 2.75, a thickness of 1 mm, and ($tan\delta$ = 0.0003). Two prototypes of open-circuited stubs were used: straight-line and meandered-line stubs. The idea behind the meandered stubs is to reduce the size of the chipless RFID tag. Spurline resonators were an alternative in [6], enabling the design of 8-bit chipless tags within a compact 60 mm × 30 mm size. The tag was fabricated on a CMET/LK4.3 substrate with a dielectric constant of 4.3 and ($tan\delta$ = 0.0018) and operates from 2.38 to 4.04 GHz. Similarly, [7] presented open-stub microstrip resonators achieving 8-bit encoding over a broad frequency range of 1.9 to 4.5 GHz. The proposed tag was fabricated on a 50 mm x 30 mm substrate with a thickness of 1.6 mm, a dielectric constant 4.4, and ($tan\delta$ = 0.0018).

Figure 1. RFID block diagram.

An 8-bit L-shaped resonator on a 20 mm × 20 mm Rogers substrate was proposed in [8], operating between 3 and 6 GHz. In [9], an 8-bit tag with discontinuities was created using embedded passive components. The tags were modified for different IDs by connecting the specified discontinuities to the transmission line (TL). In the design, eight planar capacitors were used as discontinuities. The TL length was 400 mm. Multiple resonators were used simultaneously to achieve a compact chipless RFID tag and high capacity. The resonators are equivalent to parallel and series LC circuits for different bit combinations. Further size reductions and improved frequency control were explored through bending arm resonators in [10], where two 8-bit tag prototypes used single and dual-sided configurations, operating between 4.9 GHz and 5.7 GHz. Complementary split ring resonators (CSRRs) were used in [11] to design an 8-bit 25 mm × 50 mm tag operating over a wider frequency band (3.4–7.4 GHz) on a flexible polyethylene terephthalate substrate with a dielectric constant of 3.0 and (tand ($tan\delta$ = 0.002). To achieve high-capacity encoding, a 20-bit tag using a cross-loop resonator to support polarization independent of the incident wave was presented in [12],

https://doi.org/10.31436/iiumej.v26i3.3646

operating across a wide frequency band from 3.1 to 10.1 GHz. The proposed tag was printed on a 40 mm × 40 mm RT Duroid 5880 substrate with a dielectric comprised of 2.2 ($tan\delta$ = 0.0008), and a thickness of 1 mm. In [13], a 16-bit moisture-sensing tag printed on a 13.2 mm × 19.6 mm paper substrate with (thickness 1.25 mm, dielectric constant 3.2, and ($tan\delta$ = 0.04) was designed to operate over 0.5–14 GHz. The tag was loaded with different lengths and widths of slots for tuning.

Furthermore, different shape resonators: rectangular-shaped [14], triangular-shaped [15], T-shaped [16], and C-shaped [17] were proposed. The structures in [14-16] were designed based on a finite metallic frequency selective surface (FSS) with a size of (30 mm × 30 mm), (27.5 mm × 30 mm), and (15 mm × 15 mm), respectively. In [17], a C-shaped resonator-based chipless RFID tag was proposed. Five C-shaped resonators are designed on a FR-4 substrate of size (20 mm × 40 mm). The suggested tag operates over a frequency band (2.5 and 7.5 GHz). A spiral-resonator-based chipless RFID tag was suggested in [18]. The proposed tag is printed on 108 mm × 64 mm Taconic TF-290 laminate and encodes 23 bits. The operating spectrum of the design is between 5 GHz and 10.7 GHz. In [19], a chipless RFID tag consists of (6) spiral resonators placed next to a microstrip line. The design is printed on a Taconic TLX-0 substrate and operates over a 2- and 2.5 GHz frequency band.

Chipless RFID systems often face challenges balancing miniaturization with increased bit capacity and performance over wide frequency bands. This work introduces a compact and high-capacity chipless RFID tag to achieve high capacity within a reduced size by employing back-to-back L-shaped resonators, significantly improving conventional chipless RFID designs. Two compact 8-bit chipless RFID tag prototypes based on L-shaped and U-shaped resonators are designed and analyzed to address this issue. A novel 12-bit chipless RFID tag comprising back-to-back L-shaped resonators with varying lengths, corresponding to resonant frequencies between 5 and 8.5 GHz, is proposed to increase tag capacity further.

2. CHIPLESS RFID TAG DESIGN

In this work, different prototypes of resonator-based chipless RFID tags are presented. To design an 8-bit resonator-based chipless RFID tag, the number of resonators is relative to the number of encoding bits. The resonator is designed in this work as an open-circuited length of transmission line with L-section and U-section resonator types. Such a resonator behaves as a parallel RLC resonant circuit when the length is half a wavelength ($\frac{\lambda}{2}$), or multiples of half-wavelength ($\frac{\lambda}{2}$).The input impedance of an open-circuited line of length l is, [20]:

$$Z_{in} = Z_o \frac{1+j\ tan\ \beta l\ tanh\ \alpha l}{tanh\ \alpha l + j\ tan\ \beta l} \quad (1)$$

where, Z_{in} is the input impedance of the line, Z_o is the characteristic impedance of the line, β is the propagation constant, α is the attenuation constant, and $l = \lambda/2$ at $\omega = \omega_o$. The lumped element values in the equivalent parallel RLC circuit are calculated as, [20]:

$$R = \frac{Z_o}{\alpha l} \quad (2)$$

$$C = \frac{\pi}{2\omega_o Z_o} \quad (3)$$

$$L = \frac{1}{\omega_o^2 C} \quad (4)$$

https://doi.org/10.31436/iiumej.v26i3.3646

where the resonant frequency, $\omega_o = \frac{1}{\sqrt{LC}}$. The proposed chipless RFID tag is printed on (14.5mm × 14.5mm) Rogers RT5880 substrate with a dielectric constant, $\varepsilon_r = 2.2$, and a thickness, $h = 1.575\ mm$. The effective dielectric constant, ε_{eff} , of the transmission line is calculated as, [20]:

$$\varepsilon_{eff} = \frac{\varepsilon_r+1}{2} + \frac{\varepsilon_r-1}{2}\frac{1}{\sqrt{1+12h/W}} \quad (5)$$

where, $h = 1.575\ mm$ is the thickness of the Rogers RT5880 substrate and W = 0.25 mm is the width of the resonator. The resonant frequency, f_o, is related to the resonator length, l, as:

$$f_o = \frac{c}{2l\sqrt{\varepsilon_{eff}}} \quad (6)$$

where, c is the free space velocity, l is the resonator length, and ε_{eff} is the effective permittivity. For Rogers RT5880 substrate and a transmission line of width W = 0.25 mm, the effective dielectric constant, ε_{eff} is calculated to be 1.67. The resonator length vs the corresponding resonant frequency, f_o is plotted in Matlab based on Eq.6, Figure 2. Table 1 summarizes the resonator lengths versus the corresponding resonant frequencies.

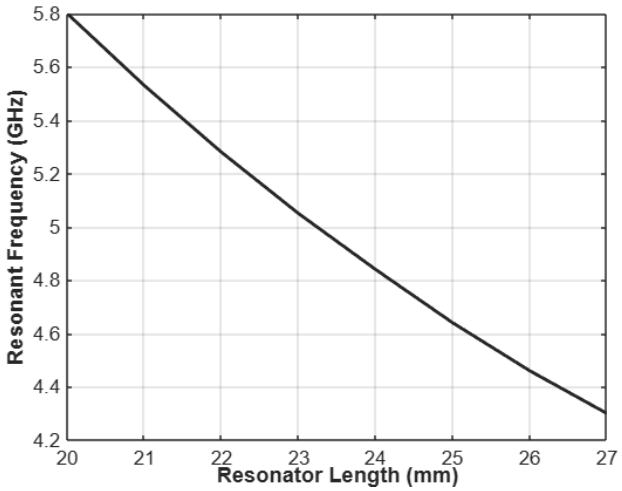


Figure 2. Resonator lengths vs corresponding resonant frequencies.

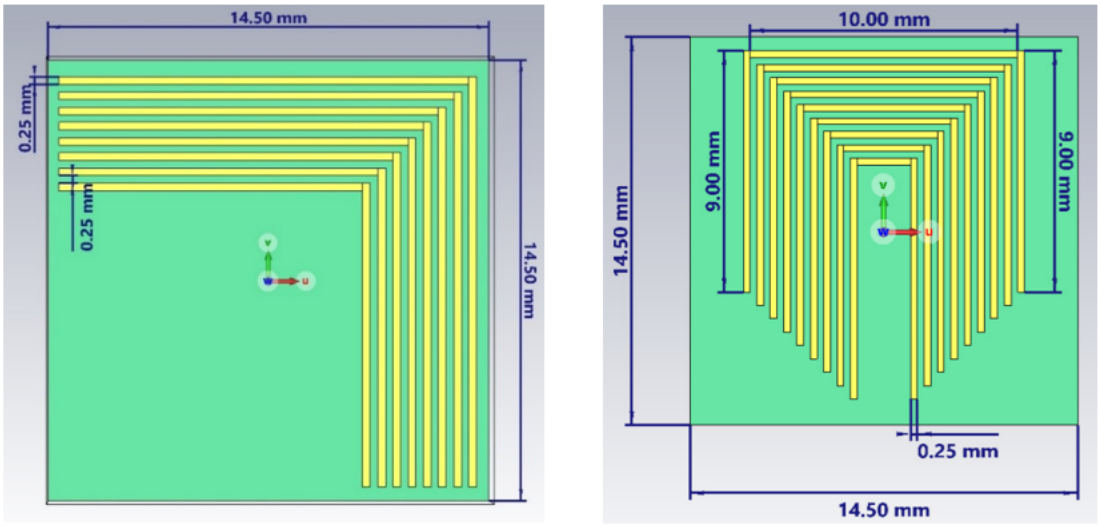
Table 1. Resonator lengths versus corresponding resonant frequencies for the 8-bit L-shaped resonator-based chipless RFID tag.

Resonator length (mm)	Resonant frequency (GHz)
27	4.30
26	4.46
25	4.64
24	4.84
23	5.05
22	5.28
21	5.53
20	5.80

The resonant frequency is inversely proportional to the resonator length. The lengths of the resonators are calculated such that the resonant frequencies extend between 4 and 6 GHz. Eight resonators are designed to correspond to an 8-bit chipless RFID tag. This work presents two compact-sized prototypes of 8-bit chipless RFID tags based on L-shaped and U-shaped resonators. The 8-bit L-shaped and U-shaped resonators are designed and simulated in CST.

https://doi.org/10.31436/iiumej.v26i3.3646

Figure 3a and b show the L-shaped and U-shaped 8-bit resonators, respectively. The proposed tags are printed on (14.5mm × 14.5mm) Rogers RT5880 substrate with a dielectric constant, $\varepsilon_r = 2.2$, and a thickness, $h = 1.575\ mm$.



(a) (b)

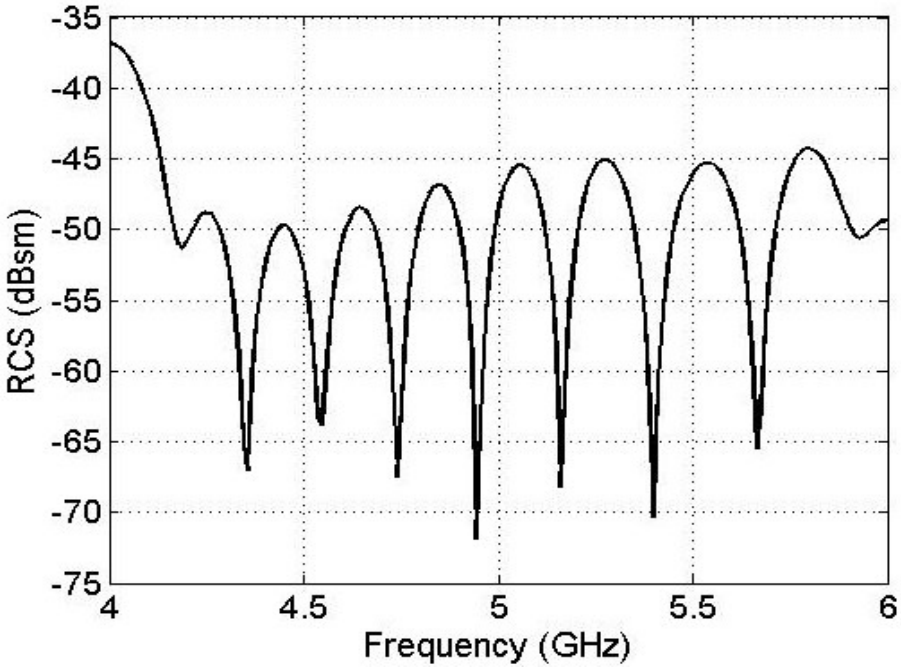
Figure 3. (a) An eight-bit resonator-based chipless RFID tag. (a) L-shaped and (b) U-shaped resonator.

The basic concept for the chipless RFID system is the same as a monostatic radar system. Where the reader radiates a power P_{rad} with a gain G, the power density incident on the tag is, [32]:

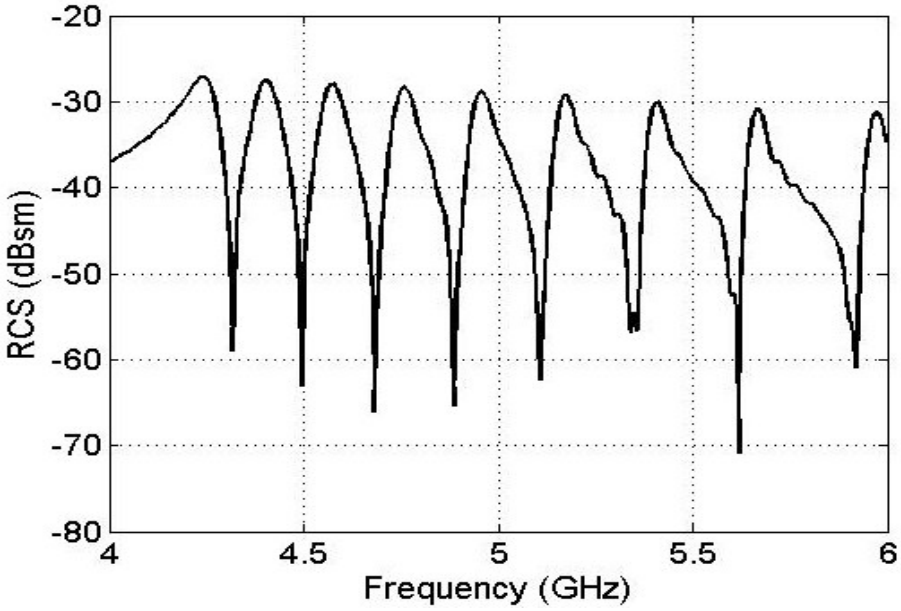
$$P_{inc} = \frac{P_{rad}G}{4\pi R^2} \tag{7}$$

where R is the distance to the tag. The tag will scatter the incident power; the ratio of the scattered power P_{scat} to the incident power P_{inc} is described as the radar cross section (RCS). It depends on the incident and reflection angles, in addition to the polarization of the incident wave. The simulated RCS responses for both the L-shaped and the U-shaped 8-bit chipless RFID tags are shown in Figure 4a and b, respectively. The resonant frequencies calculated in Table 1 using Eq.6 are slightly different from the simulated resonant frequencies obtained in Figure 4. This difference in the resonant frequencies between the calculated and simulated ones is owing to the mutual coupling effect of nearby resonators. Based on Eq.6, the resonators are considered as isolated resonators. While in CST simulation, the mutual coupling effect is considered between adjacent resonators. Due to taking into consideration the mutual coupling effect, the simulated resonant frequencies are shifted compared to the isolated resonances. The L-shaped and U-shaped resonators are chosen in this design due to their ability to create compact current paths and support multiple resonant frequencies by adjusting resonator lengths without increasing the substrate dimensions. To increase the chipless RFID tag capacity, a prototype of 12 back-to-back L-shaped resonators with different lengths corresponding to different resonant frequencies between 5 and 8.5 GHz is proposed, Figure 5. The L-shaped resonators are arranged in a back-to-back manner that supports a compact size chipless RFID tag and decreases the mutual coupling effect between the successive elements.

https://doi.org/10.31436/iiumej.v26i3.3646



(a)



(b)

Figure 4. Simulated RCS for the eight-bit resonator-based chipless RFID tag. (a) L-shaped resonator and (b) U-shaped resonator.

The proposed 12 back-to-back L-shaped resonators chipless RFID tag is printed on (15mm × 25mm) Rogers RT5880 substrate with a dielectric constant, $\varepsilon_r = 2.2$, and a thickness, $h = 1.575\ mm$. The low dielectric loss tangent (0.0009) and stable dielectric constant (2.2) over a wide frequency range make Rogers RT5880 substrate highly preferred for high-frequency RFID applications. The resonator lengths versus corresponding resonant frequencies are calculated based on Eq.6 and summarized in Table 2. The simulated RCS response for the compact size 12-bit L-shaped resonator-based chipless RFID tag is shown in Figure 6.

https://doi.org/10.31436/iiumej.v26i3.3646

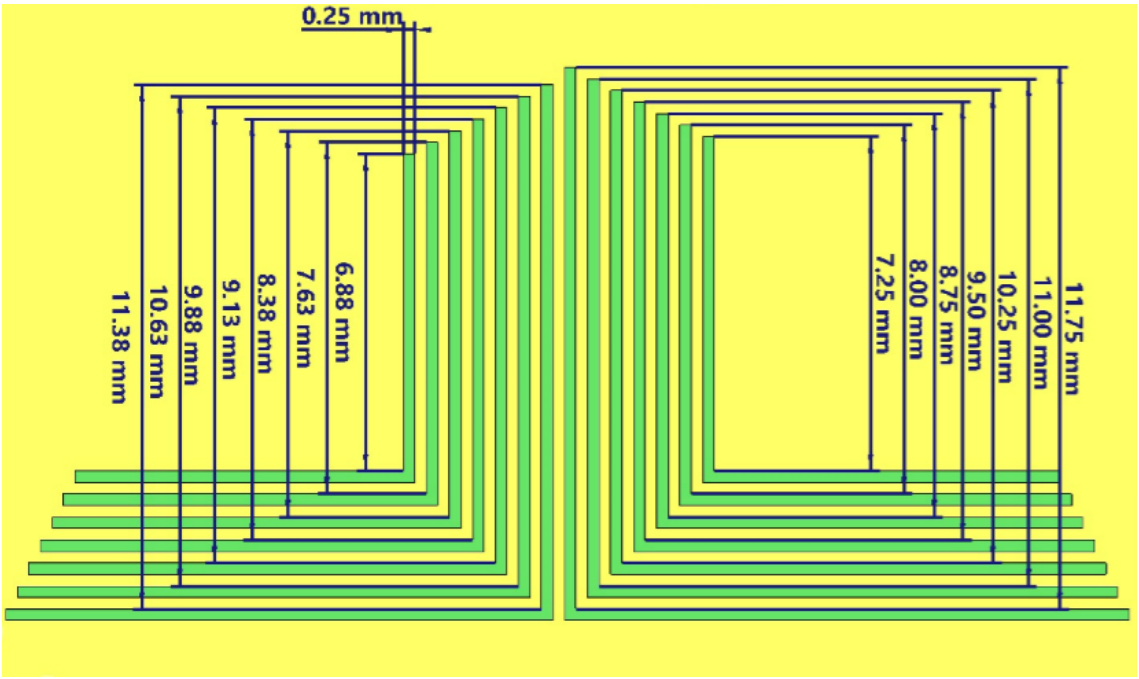


Figure 5. A compact size 12-bit L-shaped resonator-based chipless RFID tag.

The resonant frequencies calculated in Table 2 using Eq. 6 are slightly different from the simulated resonant frequencies obtained in Figure 6. This difference in the resonant frequencies between the calculated and simulated ones is due to the mutual coupling effect of nearby resonators. Based on Eq. 6, the resonators are considered as isolated resonators. While in CST simulation, the mutual coupling effect is considered between adjacent resonators. Table 3 summarizes the specifications for the three prototype chipless RFID tag designs. The suggested chipless RFID tag specifications in this study are compared to tags in the literature (Table 4). The proposed chipless RFID tag in this work offers a more compact size than those presented in references [17, 25, 27, 29, 30, 31], while achieving a higher encoding bit capacity than implementations presented in references [14, 17, 25, 27, 28, 29, 31].

Table 2. Resonator lengths versus corresponding resonant frequencies for the 12-bit L-shaped resonator-based chipless RFID tag.

Resonator length (mm)	**Resonant frequency (GHz)**
22.00	5.28
21.26	5.46
20.50	5.66
19.76	5.87
19.00	6.11
18.26	6.36
17.50	6.63
16.76	6.93
16.00	7.25
15.26	7.61
14.50	8.01
13.76	8.44

https://doi.org/10.31436/iiumej.v26i3.3646

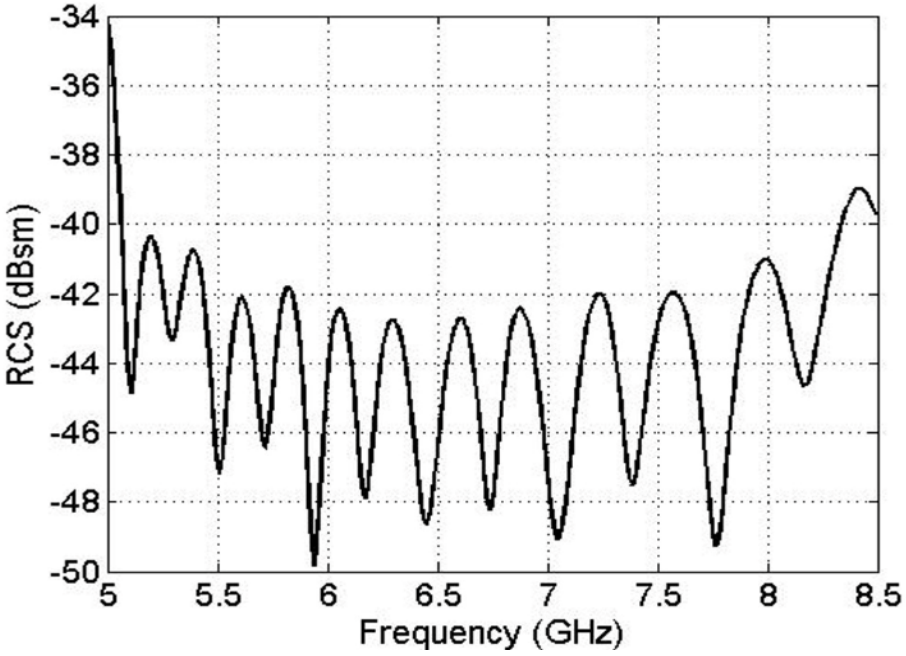


Figure 6. Simulated RCS for the compact size 12-bit L-shaped resonator-based chipless RFID tag.

Table 3. A summary of the specifications for the three prototype chipless RFID tag designs.

Chipless RFID Tag	Prototype 1	Prototype 2	Prototype 3
Structure	L-shaped resonator	U-shaped resonator	Back-to-back L-shaped resonator
Dimensions	14.5mm × 14.5mm	14.5mm × 14.5mm	15mm × 25mm
Capacity	8-bit	8-bit	12-bit
Substrate	Rogers RT5880	Rogers RT5880	Rogers RT5880
Bandwidth	4.0 to 6.0 GHz	4.0 to 6.0 GHz	5.0 to 8.5 GHz

Table 4. Comparison with other chipless RFID tags.

Benchmark	Capacity (bit)	Bandwidth (GHz)	Substrate	Dimensions (mm^2)
[8]	8	3.0 – 6.0	Rogers RT5880	20 mm × 20 mm
[11]	8	3.4 – 7.4	Polyethylene terephthalate	25 mm × 50 mm
[14]	5	2.0 – 8.0	FR-4	30 mm × 30 mm
[15]	10	4.0 – 11.0	FR-4	27.5 mm × 30 mm
[16]	10	5.0 – 15.0	FR-4	15 mm × 15 mm
[17]	5	2.5 – 7.5	FR-4	20 mm × 40 mm
[18]	23	5.0 – 10.7	Taconic TF-290 laminate	108 mm × 64 mm
[19]	6	2.0 – 2.5	Taconic TLX-0	66 mm × 60 mm
This work	12	5.0 – 8.5	Rogers RT5880	15 mm × 25 mm

3. CONCLUSION

Miniaturized L-shaped and U-shaped resonator-based 8-bit Chipless RFID tags are presented. The proposed tags are printed on (14.5mm × 14.5mm) Rogers RT5880 substrate with a 4 to 6 GHz frequency band. To increase the chipless RFID tag capacity, a prototype of 12 back-to-back L-shaped resonators with different lengths corresponding to different resonant frequencies between 5 and 8.5 GHz is proposed. The proposed 12 back-to-back L-shaped resonators chipless RFID tag is printed on (15mm × 25mm) Rogers RT5880 substrate with a

frequency band from 5 GHz to 8.5 GHz. The simulated RCS responses for the compact size L-shaped and U-shaped resonator-based chipless RFID tags are calculated. The resonant frequencies for isolated resonators are calculated and compared with those of the mutually coupled resonators. Due to current limitations in fabrication, the manufacturing and experimental validation of the proposed tag are considered part of our future work to verify the simulation results. Additionally, future efforts will focus on using flexible substrates for applications on wearable or curved surfaces.

REFERENCES

[1] S. Preradovic and N. C. Karmakar, (2010), Chipless RFID: Bar Code of the Future. IEEE Microwave Magazine, 11(7), 87–97. DOI:10.1109/MMM.2010.93857.

[2] H. F. Huang and L. Su, (2017), A Compact Dual-Polarized Chipless RFID Tag by Using Nested Concentric Square Loops. IEEE Antennas and Wireless Propagation Letters, 16, 1036–1039. DOI:10.1109/lawp.2016.2618928.

[3] S. Preradovic, N.C. Karmakar, (2012), Low cost chipless RFID systems, in Multiresonator-Based Chipless RFID, Springer, New York, pp. 9–24. DOI:10.1007/978-1-4614-2095-8.

[4] I. Jalaly, I.D. Robertson, (2005), Capacitively-tuned split microstrip resonators for RFID barcodes. In IEEE European Microwave Conference, vol. 2. DOI:10.1109/EUMC.2005.1610138

[5] M. E. Jalil, M.K.A. Rahim, N.A. Samsuri, R. Dewan, (2014), Chipless RFID tag based on meandered line resonator. In IEEE Asia Pacific Conference on Applied Electromagnetics (APACE), Malaysia, pp. 203–206. DOI: 10.1109/APACE.2014.7043780.

[6] M. Sumi, R. Dinesh, C.M. Nijas, S. Mridula, P. Mohanan, (2014), Frequency coded chipless RFID tag using spurline resonators. Radio engineering, vol. 23, no. 1, 203–208.

[7] C. M. Nijas, R. Dinesh, U. Deepak, A. Rasheed, S. Mridula, K. Vasudevan, P. Mohanan, (2012), Chipless RFID Tag Using Multiple Microstrip Open Stub Resonators. IEEE Transactions on Antennas and Propagation, vol. 60, no. 9, pp.4429-4432. DOI:10.1109/TAP.2012.2207081.

[8] M. S. Hashimi, V. Sharma, (2020), Design, analysis, and realisation of chipless RFID tag for orientation independent configurations. The Journal of Engineering, vol. 2020 no. 5, pp. 189-196. Doi: 10.1049/joe.2019.0920.

[9] L. Zheng, R. Saul, L. Zhang, B. Shao, L. R. Zheng, (2008), Design and implementation of a fully reconfigurable chipless RFID tag using Inkjet printing technology. 2008 IEEE International Symposium on Circuits and Systems (ISCAS), May 2008, USA. DOI: 110.1109/ISCAS.2008.4541720.

[10] S. Jia; L. Xiuping; Z. Hua, (2017), Multiresonator-based chipless RFID system for low-cost application. 2017 Progress in Electromagnetics Research Symposium - Fall (PIERS - FALL), Nov. 2017, Singapore. DOI: 10.1109/PIERS-FALL.2017.8293197.

[11] M. E. Jalil, M. K. A. Rahim, N. A. Samsuri, R. Dewan, (2016), Flexible printed chipless RFID tag using metamaterial–split ring resonator. Applied Physics A Materials Science and Processing, (2016) 122:348. DOI 10.1007/s00339-016-9865-5.

[12] V. R. Sajitha, C. M. Nijas, T. K. Roshna, Kesavath Vasudevan, P. Mohanan, (2016), Compact cross loop resonator based chipless RFID tag with polarization insensitivity. Microwave and Optical Technology Letters, vol. 58, no. 4, pp. 944–947. Doi.org/10.1002/mop.29706.

[13] N. Javed, A. Habib, A. Akram, Y. Amin, and H. Tenhunen, (2016), 16-bit frequency signatured directly printable tag for organic electronics. IEICE Electronics Express, vol. 13, no. 11. DOI:10.1587/elex.13.20160406.

[14] F. Costa, S. Genovesi, A. Monorchio, (2013), A chipless RFID based on multiresonant high impedance surfaces. IEEE Transactions on Microwave Theory and Techniques, vol. 61, no. 1, pp.146-153. DOI: 10.1109/ TMTT.2012.2227777.

[15] S. Rauf, M. A. Riaz, H. Shahid, M. S. Iqbal, Y. Amin, H. Tenhunen, (2017), Triangular loop resonator based compact chipless RFID tag. IEICE Electronic Express, vol. 14, no. 4. DOI: 10.1587/elex.14. 20161262.

[16] M. A. Riaz, H. Shahid, S. Z. Aslam, Y. Amin, A. Akram, H. Tenhunen, (2017), Novel T-shaped resonator based chipless RFID tag. IEICE Electronic Express, vol. 14, no.18. DOI: 10.1587/elex.14.20170728.

[17] A. Vena, E. Perret, S. Tedjini, (2011), Chipless RFID Tag Using Hybrid Coding Technique. IEEE Transactions on Microwave Theory and Techniques, vol. 59, no. 12, pp. 3356 - 3364. DOI: 10.1109/TMTT.2011.2171001.

[18] S. Preradovic, S. Roy, and N. C. Karmakar, (2009), Fully printable multi-bit chipless RFID transponder on flexible laminate. In Proc. Asia-Pacific Micro. Conf., Singapore, pp. 2371–2374. DOI: 10.1109/APMC.2009.5385460.

[19] S. Preradovic, I. Balbin, N. C. Karmakar, and G. F. Swiegers, (2009), Multiresonator-based chipless RFID system for low-cost item tracking. IEEE Trans. Microw. Theory Tech., vol. 57, no. 5, pp. 1411–1419. DOI: 10.1109/RFID.2008.4519383.

[20] Pozar, D.M. (2005). Microwave engineering (2nd ed.). USA: John Wiley and Sons Inc.

https://doi.org/10.31436/iiumej.v26i3.3652

LORA-DRIVEN ALGORITHMS FOR ACCURATE RETURN-TO-HOME (RTH) PREDICTION IN UNMANNED AERIAL VEHICLES

MOHAMAD HAZWAN MOHD GHAZALI[1], HAN HENG[2], IZZATI SALEH [2], WAN RAHIMAN[2*]

[1]*Department of Mechanical and Manufacturing Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia*
[2]*School of Electrical and Electronic Engineering, Universiti Sains Malaysia Engineering Campus, 14300 Nibong Tebal, Penang, Malaysia*

**Corresponding author: wanrahiman@usm.my*

(Received: 17 March 2025; Accepted: 23 June 2025; Published online: 9 September 2025)

ABSTRACT: Unmanned aerial vehicles (UAVs) are prone to crashes due to user inexperience, technical issues, or adverse weather conditions. Most commercial UAVs have a real-time monitoring feature, where important UAV parameters can be monitored. However, low-cost, non-ready-to-fly UAVs often lack these capabilities, making assessing their condition challenging or taking preventive measures when faults occur. This paper presents a novel return-to-home (RTH) prediction algorithm for UAVs, which is triggered when a fault is detected in the UAV. The key contribution of this study is the innovative application of LoRa data in the design of the prediction algorithm, which sets it apart from existing reviews and research that have not explored this approach. The data from the LoRa wireless communication network will be utilized in this algorithm, which consists of three critical parameters: the speed of the UAV (V), the flight range (R), and the battery level (T). Specifically, the algorithm utilizes LoRa's received signal strength indicator (RSSI) data to estimate the flight range and is designed for use within a 1 km flight radius. It provides actionable recommendations to UAV pilots to return the UAV to its home position or to land immediately, which can be accessed through a mobile application. This approach enhances safety by reducing the risk of UAV crashes and ensuring timely interventions.

ABSTRAK: Kenderaan udara tanpa pemandu (UAV) terdedah kepada kemalangan kerana faktor seperti pengguna kurang berpengalaman, isu teknikal atau keadaan cuaca buruk. Kebanyakan UAV komersial yang terdapat di pasaran didatangi dengan ciri pemantauan masa sebenar, di mana parameter penting UAV boleh dipantau. Walau bagaimanapun, UAV berkos rendah dan tidak sedia terbang sering kekurangan keupayaan ini, menjadikannya mencabar dalam menilai keadaan atau mengambil langkah pencegahan apabila berlaku kerosakan. Kajian ini mengkaji algoritma baharu ramalan pulang ke pangkalan (RTH) untuk UAV, tercetus apabila kerosakan dikesan dalam UAV. Sumbangan utama kajian ini adalah aplikasi inovatif data LoRa dalam reka bentuk algoritma ramalan, membezakannya daripada kajian terdahulu yang belum menggunakan LoRa dalam konteks ini. Data dari rangkaian komunikasi tanpa wayar LoRa digunakan dalam algoritma ini, iaitu terdiri daripada tiga parameter penting: kelajuan UAV (V), penerbangan (R), dan paras bateri (T). Secara khusus, algoritma menggunakan data penunjuk kekuatan isyarat (RSSI) yang diterima LoRa bagi menganggarkan julat penerbangan dan direka bentuk bagi kegunaan radius penerbangan 1 km. Kajian ini memberi cadangan tindakan yang boleh diambil oleh juruterbang UAV, sama ada bagi mengembalikan UAV ke kedudukan asal atau mendarat serta-merta, yang boleh diakses melalui aplikasi mudah alih. Pendekatan ini meningkatkan keselamatan dengan

mengurangkan risiko kemalangan UAV dan memastikan campur tangan pengguna tepat pada masanya.

KEYWORDS: *Unmanned Aerial Vehicle, Wireless Communication, Predictive Algorithms, Flight Time, Signal Strength*

1. INTRODUCTION

Drones, also known as unmanned aerial vehicles (UAVs), have undergone remarkable advancements over the past fifteen years [1,2,3]. While initially limited to military applications, UAVs are now widely available for civilian use. Their growing popularity stems from affordability and versatility, enabling their use across various applications. Typically, a UAV is equipped with a global positioning system (GPS) and operates via a radio channel transmitter and receiver. One of the most appealing aspects of UAVs is their adaptability, as users can customize them by incorporating additional hardware and implementing specific algorithms to meet their needs.

UAVs can be categorized into two main types: fixed-wing and multirotor [4]. Fixed-wing UAVs feature a pair of wings that passively generate lift as the UAV moves through the air at a specific angle, while multirotors rely on the speed and direction of their motors for movement. Multirotors typically have motors, propellers, and electronic speed controllers (ESCs). Multirotors have lower flight speeds, shorter ranges, and reduced flight durations than fixed-wing UAVs due to the substantial power required to generate lift and remain airborne [5]. Some of the well-known UAVs include the Parrot Disco (fixed-wing), DJI Phantom 4 (multirotor), and DJI Matrice (multirotor). However, UAVs, particularly multirotors, are susceptible to crashes, potentially causing unwanted casualties. These incidents can arise from factors such as communication interference, battery issues, extreme weather conditions, and failures in mechanical or electrical components.

Several fault detection and diagnosis systems have been developed to monitor the condition of UAVs, based on vibration, acoustic, and current parameters, to address these challenges. Al-Haddad et al. [6] proposed a UAV unbalanced fault classification approach by analyzing the frequency-domain vibration signals. An ADXL335 accelerometer was applied to acquire the vibration data on the quadcopter in the indoor laboratory environment. Pourpanah et al. [7] analyzed the vibration characteristics to differentiate between healthy and faulty propellers (normal, 5%, 10%, or 15% damaged) of a quadcopter, where four accelerometers are mounted at the bottom of each motor. In the study conducted by Altinors et al. [8], they proposed a method for detecting faults in drone motors using acoustic data. They employed machine learning techniques and achieved an accuracy of over 99% in detecting motor failures. Lee et al. [9] proposed a method for diagnosing faults in hexacopter motors using an IR sensor to compute angular speed and measure current at the ESC input. Next, Iannace et al. [10] utilized the noise generated by the drone to develop a classification model based on neural networks to identify broken propeller blades. However, the previously mentioned studies do not address countermeasures for faults occurring during drone operations. While most research on return-to-home (RTH) prediction focuses on UAV navigation or path planning [11-14], there is a limited exploration of the predictive aspect of determining whether a drone can successfully return to its home position. Additionally, no studies have proposed RTH prediction based on LoRa wireless communication data.

This study introduces a novel RTH prediction algorithm for UAVs based on LoRa wireless communication data. When a fault is detected during flight, the algorithm evaluates whether the UAV can safely return to its home position or must be landed immediately. While this

https://doi.org/10.31436/iiumej.v26i3.3652

feature is available in most commercial UAVs, it remains absent in certain drones, particularly non-ready-to-fly models. The research contributions are as follows:

i) Novel RTH prediction algorithm based on LoRa communication data

ii) Enhanced UAV safety, especially for custom-built drones

The remainder of the paper is organized as follows: Section 2 presents the LoRa wireless communication system and data collection. Next, Section 3 discusses the proposed RTH prediction algorithm, which consists of three parameters: flight range, battery level, and UAV speed. Finally, Section 4 concludes the study.

2. LORA WIRELESS COMMUNICATION

LoRa is one of the low-power wide-area network (LPWAN) technologies from the LoRa Alliance. It attracts considerable attention because it can efficiently trade communication ranges with big data rates, thus allowing IoT applications at an urban scale [15]. LoRa-based communication networks have been implemented in several applications such as precision agriculture [16], traffic monitoring [17], rescue operations [18], and military [19]. LoRa is based on the chirp spread spectrum modulation, and its spreading factor and coding rate parameters can be adjusted to achieve the desired communication network [20]. Depending on the operational region, LoRa mainly operates in the unlicensed ISM bands at 433, 868, or 915 MHz. In Malaysia, 915 MHz is a permitted free band for LoRa applications according to the Malaysian Communications and Multimedia Commission (MCMC). In this research, the Cytron LoRa radio frequency modulation (RFM) shields were utilized as a transmitter and receiver equipped with 2 dBi antennas, as shown in Figure 1. The Cytron LoRa Shield stacks directly onto an Arduino UNO and communicates via SPI, enabling the Arduino to send and receive data through the LoRa transceiver. It generally supports point-to-point (P2P) communication, while LoRaWAN requires additional configuration and infrastructure. The transmit power of both LoRa shields was set at 23 dBm, and a 915 MHz frequency band was utilized in this research. The technical specifications of the Cytron LoRa shield are listed in Table 1.

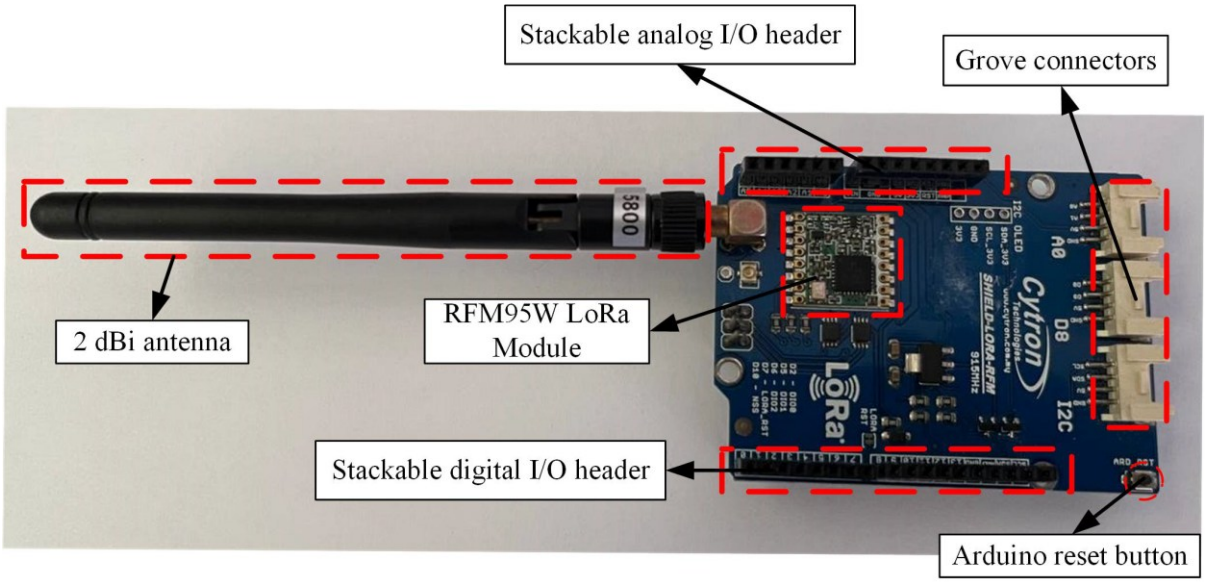


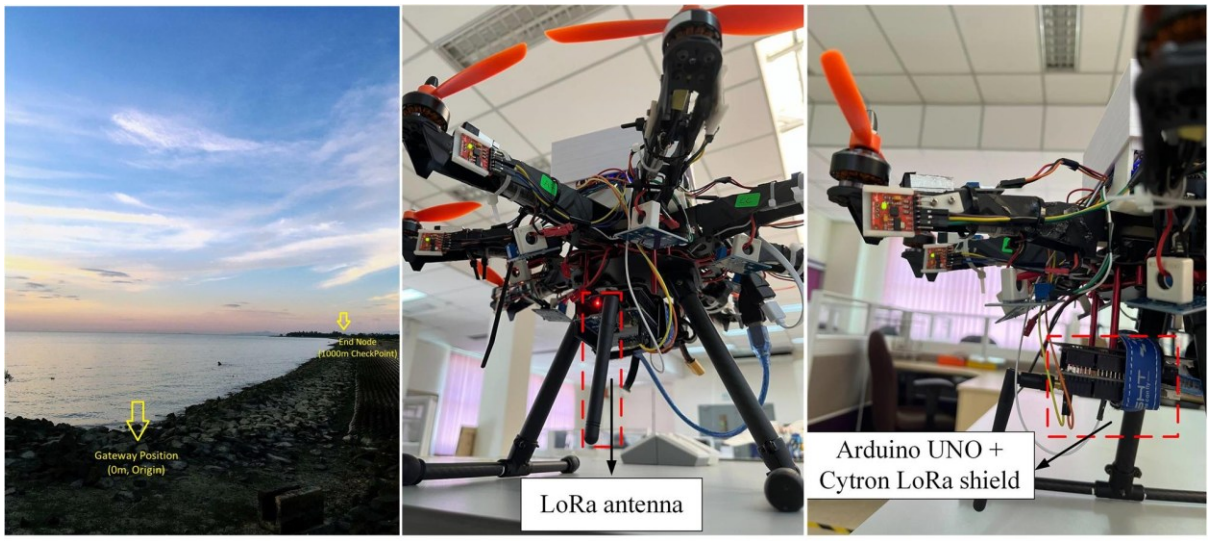
Figure 1. Hardware structure of Cytron LoRa shield

https://doi.org/10.31436/iiumej.v26i3.3652

Table 1. Technical specification of Cytron LoRa shield

No.	Specification	Information
1	Dimension	79 mm x 53 mm x 15 mm (without antenna)
2	Weight	30 g
3	Frequency Band	915 MHz
4	High Receiver Sensitivity	Down to -146 dBm
5	Transmit Power	Adjustable up to +14 dBm
6	Antenna	2 dBi antenna

The LoRa data collection was conducted in Ban Pecah, located in Parit Buntar, Perak, Malaysia, as depicted in Figure 2(a). This location was chosen because it can cater to the 1 km direct line-of-sight feature, our maximum communication range set in this research. The 1 km range is selected because, at this range, the user cannot determine most of the UAV's behaviour, even the large ones, through the naked eye. Also, if no FPV is implemented, given small UAVs, it is hard to inspect the UAV's behaviour even at 500 m visually. Although the UAV flying range and LoRa applicability are more than 1 km, only the 1 km range is tested for the sake of the experiment. As shown in Figure 2(b), the LoRa transmitter is installed at the bottom part of the UAV body. In contrast, the LoRa receiver is attached to the personal computer for real-time data recording. Data collection begins at a communication range of 100 m and increments by 100 m with each step, continuing until the final range of 1 km, as summarized in Table 2. The algorithm utilizes LoRa's received signal strength indicator (RSSI), measured in decibels (dBm). Theoretically, the closer the communication range, the higher the RSSI value. The Storm Drone 8 multirotor UAV is adopted in this study, as shown in Figure 2(b).



(a) (b)

Figure 2. (a) LoRa data collection site in Ban Pecah, Perak, Malaysia and (b) Mounting location of Cytron LoRa shield at the UAV

https://doi.org/10.31436/iiumej.v26i3.3652

Table 2. LoRa's RSSI recorded at the communication range of up to 1 km across five trials

Communication Range (m)	RSSI (dBm)				
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
100	-66 to -68	-67 to -69	-66 to -68	-66 to -68	-67 to -68
200	-71 to -73	-71 to -72	-70 to -72	-71 to -72	-70 to -72
300	-73 to -76	-73 to -76	-73 to -75	-73 to -76	-74 to -76
400	-81 to -83	-82 to -83	-82 to -84	-82 to -83	-82 to -83
500	-83 to -85	-83 to -85	-83 to -84	-83 to -85	-83 to -85
600	-84 to -86	-84 to -86	-84 to -86	-84 to -86	-84 to -86
700	-87 to -90	-87 to -89	-88 to -90	-87 to -90	-88 to -90
800	-87 to -90	-87 to -90	-87 to -90	-87 to -90	-87 to -90
900	-89 to -92	-90 to -92	-89 to -92	-90 to -92	-90 to -92
1000	-91 to -93	-91 to -93	-91 to -93	-91 to -93	-92 to -93

3. UAV RETURN TO HOME PREDICTION ALGORITHMS

The RTH prediction algorithm is computed to assess whether bringing the UAV back to its initial location is feasible. If the algorithm determines it is not possible, it's recommended that the UAV be landed immediately. The RTH algorithm is only executed when the faulty output is received from the Arduino microcontroller. This algorithm takes into account three parameters: the speed of the UAV (V), the flight range (R), and the battery level (T). These parameters are readily obtainable from most commercial UAVs, but certain UAVs, such as Quadcopter F450 and Hexacopter F550, lack the necessary instruments to determine them.

The maximum flying speed of Storm Drone 8 used is experimentally determined to be 13 m/s. The battery level of the UAV is determined using flight time data recorded from the Arduino microcontroller, which begins at the take-off stage. Thus, based on the flight time data obtained, the battery level of the UAV can be predicted. The flight range is obtained based on LoRa's RSSI data. For instance, if the RSSI obtained is less than -95 dBm, the flight range is determined to be approximately 1 km. The LoRa's RSSI data are recorded in a direct line-of-sight environment, as discussed in Section 2.0.

To acquire flight time data for fault cases, the motor failure conditions were simulated by disconnecting the ESC wires (red and black) from the power supply, as illustrated in Figure 3. During flight time data recording, the UAV hovers at a maximum altitude of 2 m. The verdict on whether the UAV can perform return-to-home is computed by:

$$RTH = T + RT \tag{1}$$

where RTH is the return time, R is the flight range data, V is the flying speed of the UAV, and T is the battery level obtained from the flight time data. The value obtained is then compared with the assumed maximum flight time of the UAV, which in this case is approximately 480 seconds, as listed in Table 3. The maximum flight time is assumed to be 480 seconds by considering the flight time of the minimum partial safe configurations, which is the failure of two motors. Hence, if the flight time is underestimated, the probability of the UAV crashing upon its return home is diminished. Referring to Table 3, the flight tests revealed that the flight time decreased to approximately 490 to 500 seconds, representing a 12.21% reduction compared to the UAV's flight time under a healthy condition. The decrease in the flight time is due to the much higher power required to generate enough thrust force during hovering. Conditions involving the failure of two adjacent motors or more than two motors are not recorded, as the UAV cannot hover and experiences an immediate crash upon takeoff.

https://doi.org/10.31436/iiumej.v26i3.3652

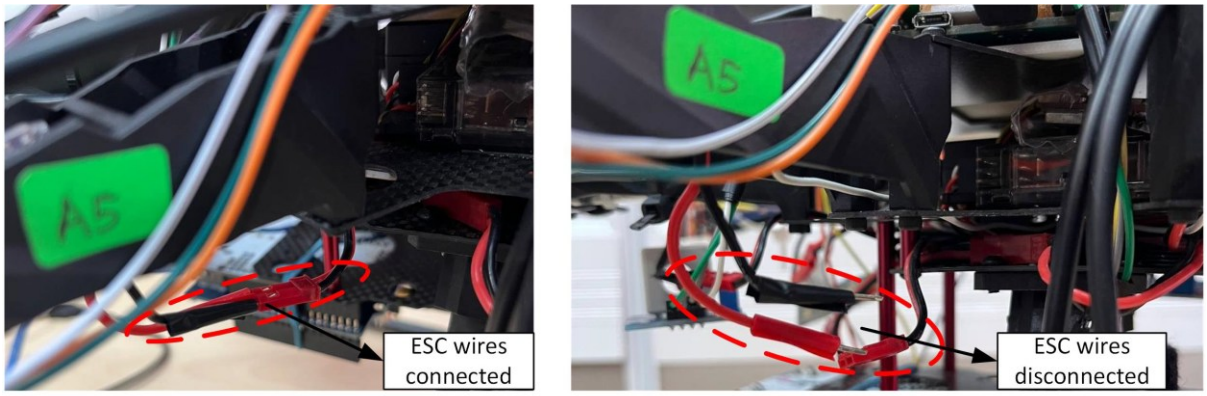
Table 3. Flight time recorded at different fault configurations

Trial	Motor 1	Motor 2	Motor 3	Moor 4	Motor 5	Motor 6	Motor 7	Motor 8	Flight Time (Sec)
1	Healthy	Healthy	Healthy	Healthy	Healthy	Healthy	Healthy	Healthy	565
2	Healthy	Healthy	Healthy	Healthy	Healthy	Healthy	Faulty	Healthy	550
3	Healthy	Healthy	Faulty	Healthy	Healthy	Healthy	Healthy	Healthy	550
4	Healthy	Healthy	Faulty	Healthy	Healthy	Healthy	Faulty	Healthy	500
5	Healthy	Healthy	Healthy	Healthy	Faulty	Healthy	Healthy	Healthy	554
6	Faulty	Healthy	Healthy	Healthy	Healthy	Healthy	Healthy	Healthy	552
7	Faulty	Healthy	Healthy	Healthy	Faulty	Healthy	Healthy	Healthy	496
8	Healthy	Faulty	Healthy	Healthy	Healthy	Faulty	Healthy	Healthy	492
9	Healthy	Faulty	Healthy	Healthy	Healthy	Healthy	Healthy	Healthy	554
10	Healthy	Healthy	Healthy	Healthy	Healthy	Faulty	Healthy	Healthy	551
11	Healthy	Healthy	Healthy	Faulty	Healthy	Healthy	Healthy	Faulty	495
12	Healthy	Healthy	Healthy	Faulty	Healthy	Healthy	Healthy	Healthy	551
13	Healthy	Healthy	Healthy	Healthy	Healthy	Healthy	Healthy	Faulty	550

The proposed RTH prediction algorithm is presented as Algorithm 1 and can be observed in Fig. 4. The algorithm starts by computing the UAV's flying speed and battery level. Then, depending on the performance output (PO) value, the proposed algorithm is triggered. We set the PO threshold to be greater than 5 for the algorithm to activate, which indicates that the UAV is behaving abnormally. Once triggered, the algorithm utilizes the UAV's flight time. The flight range can be determined based on LoRa wireless communication data. Subsequently, the return time and return-to-home (RTH) value are computed according to Eq. [1]. If the RTH value exceeds 480, the UAV can safely return to the home position. However, if the value

https://doi.org/10.31436/iiumej.v26i3.3652

exceeds 480, it indicates insufficient time for the UAV to return home and it must land immediately.



(a) (b)

Figure 3. The UAV motor configurations involved in the experiment: (a) Healthy motor and (b) Condition that represents a faulty motor

ALGORITHM 1 Proposed RTH Prediction Algorithm

```
FLYING SPEED (V = 13 m/s)
BATTERY LEVEL (T)
        if (PO < 5)
        setMillis(newvalue)
        T = millis() / 1000
FLIGHT RANGE (R)
        if(rf 95.lastRssi() >= −40)
        R = 5
        if(rf 95.lastRssi() < −40) and (rf 95.lastRssi() >= −45)
        R = 10
        if(rf 95.lastRssi() < −45) and (rf 95.lastRssi() > −60)
        R = 80
        if(rf 95.lastRssi() <= −60) and (rf 95.lastRssi() >= −70)
        R = 100
        .
        .
        .
        else if(rf 95.lastRssi() < −95)
        R = 1000
RETURN TIME (RT = T/V)
PREDICT RETURN-TO-HOME (RTH = T+RT)
        if(RTH < 480)
        Return to home
        if(RTH > 480)
        Land immediately
```

https://doi.org/10.31436/iiumej.v26i3.3652

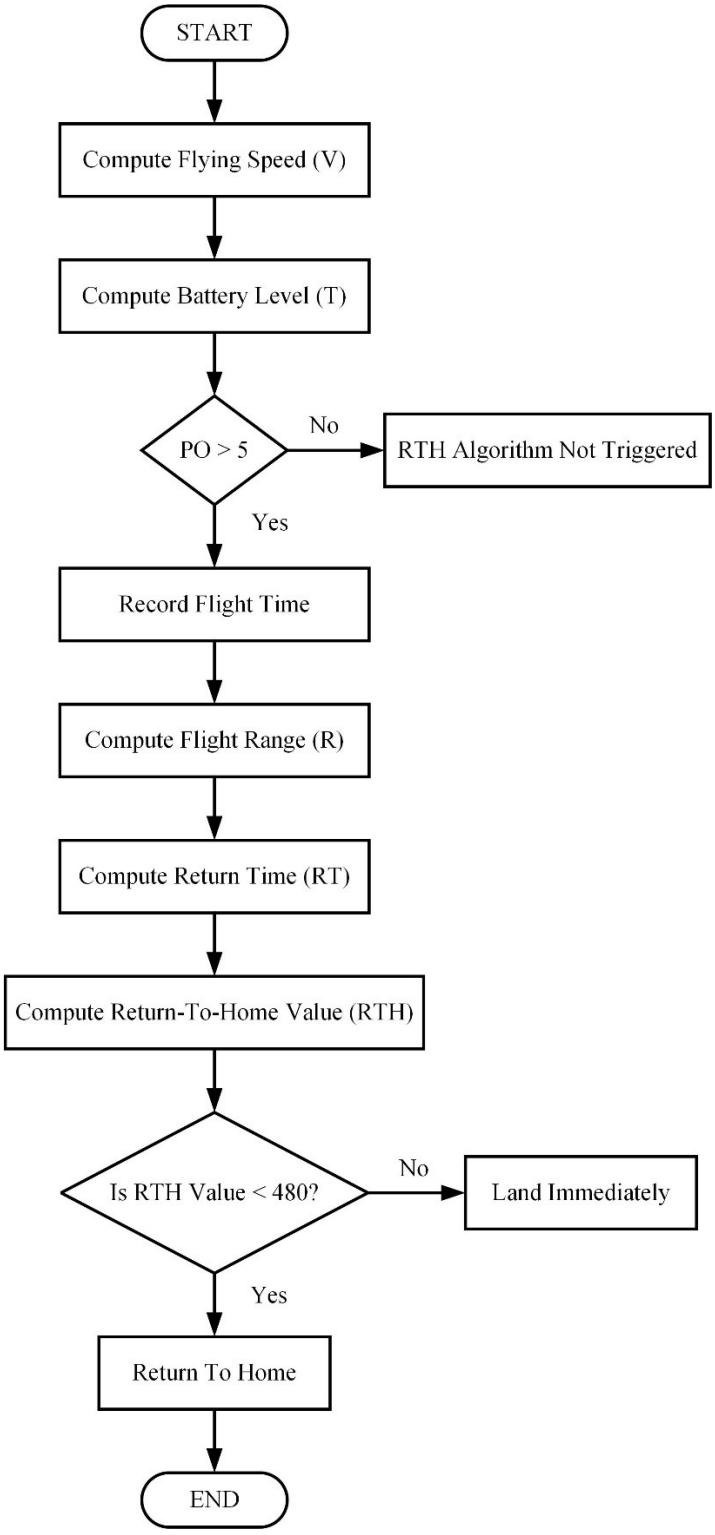


Figure 4. Flowchart of the proposed RTH algorithm

Table 4 shows the latency test for the proposed method conducted again in Ban Pecah, Parit Buntar, Perak, Malaysia. This location was chosen because it can cater to the 1 km direct line-of-sight feature, our maximum communication range set in this research, as discussed in Section 2.0. Based on Table 4, the output of the proposed algorithm can be determined in approximately 1 second within a communication range of up to 300 m, without any delay. Nonetheless, when the communication range exceeds 500 m, delays of up to four seconds are

observed. Compared to the work conducted by Fadheli et al. [13], the proposed method does not require a camera or image processing techniques, as the pilot performs the RTH process. Although their approach automates the RTH process, our proposed method focuses on predicting the feasibility of the RTH process.

Table 4. Latency tests conducted in a 1 km direct line-of-sight communication range

Trial	Range [m]	Time Taken for Decision-Making [s]	Delay [s]
1	1	1	0
2	100	1	0
3	200	1	0
4	300	1	0
5	500	1.5	0.5 (max)
6	700	3	2 (max)
7	1000	5	4 (max)

4. CONCLUSION

This study introduced a novel RTH prediction algorithm for UAVs, utilizing LoRa wireless communication data to assess whether a UAV can safely return to its home position or should be landed immediately. By integrating LoRa's RSSI data with parameters such as UAV speed and battery level, the algorithm offers a new method for enhancing UAV safety and operational reliability, specifically targeting UAVs with no RTH feature. The experimental results demonstrated that the failure of two opposite motors can reduce the UAV flight time by up to 60 seconds, where this data will be the benchmark for the proposed algorithm to make a prediction. The findings highlight a new way of incorporating LoRa technology into UAV systems to mitigate crash risks by predicting the UAV's ability to return to the home position. Additionally, the LoRa wireless communication technology demonstrated in this study can be adapted for various applications, including environmental monitoring and communication relay, simultaneously enhancing its versatility. The proposed work is limited to a 1km flight range and direct line-of-sight conditions. Future work may focus on expanding the algorithm's applicability to non-line-of-sight environments and further integrating additional sensor data to enhance predictive accuracy. Additionally, a statistical comparison with algorithms from previous works will be conducted. This contribution marks a significant advancement in enhancing UAV operational safety and adds new knowledge to the domain of fault mitigation by utilizing LoRa data to prevent drone crashes. The proposed approach can also be integrated with the existing UAV fault detection and diagnosis systems available in the literature.

ACKNOWLEDGEMENT

We would like to thank Synvue Sdn Bhd (1236749-D) for insights and feedback on the project development.

REFERENCES

[1] Barbedo JGA. (2019) A review on the use of unmanned aerial vehicles and imaging sensors for monitoring and assessing plant stresses, Drones, 3(2):40-67.

[2] Mohsan SAH, Othman NQH, Li Y, Alsharif MH, Khan MA. (2023) Unmanned aerial vehicles (UAVs): Practical aspects, applications, open challenges, security issues, and future trends, Intelligent Service Robotics, 16(1):109-137.

[3] Harun MH, Abdullah SS, Aras MSM, Bahar MB, Ali@Ibrahim F. (2023) 3D Collision avoidance system for unmanned aerial vehicle (UAV) with decentralized approach, IIUM Engineering Journal, 24(2):373-390.

[4] Boon MA, Drijfhout AP, Tesfamichael S. (2017) Comparison of a fixed-wing and multi-rotor UAV for environmental mapping applications: A case study, In International Conference on Unmanned Aerial Vehicles in Geomatics, Bonn, Germany.

[5] Zufferey R, Siddall R, Armanini S, Kovac M. (2022) Synthetic aerial aquatic locomotion. In Between Sea and Sky: Aerial Aquatic Locomotion in Miniature Robots, Springer, 2022:33-41.

[6] Al-Haddad LA, Jabe A.A, Neranon P, Al-Haddad SA. (2023) Investigation of frequency-domain-based vibration signal analysis for UAV unbalance fault classification, Engineering and Technology Journal, 41(7):915-923.

[7] Pourpanah F, Zhang B, Ma R, Hao Q. (2018) Anomaly detection and condition monitoring of UAV motors and propellers, In 2018 IEEE Sensors, New Delhi, India.

[8] Altinors A, Yol F, Yaman O. (2021) A sound based method for fault detection with statistical feature extraction in UAV motors, Applied Acoustics, 183:1-14.

[9] Lee JY, Lee WT, Ko SH, Oh HS. (2020) Fault classification and diagnosis of UAV motor based on estimated nonlinear parameter of steady-state model, International Journal of Mechanical Engineering and Robotics Research, 10(1):22-31.

[10] Iannace G, Ciaburro G, Trematerra A. (2019) Fault diagnosis for UAV blades using artificial neural network, Robotics, 8(3):59-76.

[11] Zhang W, Ma X, Hu W. (2024) Research on dynamic return method of UAV formation based on machine learning. In 9th Asia-Pacific Conference on Intelligent Robot Systems, Dalian, China.

[12] Alkhatib M, Nayfeh M, Al Shamaileh K, Kaabouch N, Devabhaktuni V. (2025) A return-to-home unmanned aerial vehicle navigation solution in global positioning system denied environments via bidirectional long short-term memory reverse flightpath prediction, Engineering Applications of Artificial Intelligence, 140.

[13] Fadheli H, Khalid SKA, Fadzil LM. (2025) A non-GPS return to home algorithm for drones using convolutional neural network, Journal of Advanced Research in Applied Sciences and Engineering Technology, 51(1):15-27.

[14] Hung I, Unger D, Zhang Y, Kulhavy D. (2024) Assessing drone return-to-home landing accuracy in a woodland landscape, Drones and Autonomous Vehicles, 1(3):1-13.

[15] Ghazali MHM, Teoh K, Rahiman W. (2021) A systematic review of real-time deployments of UAV-based LoRa communication network, IEEE Access, 9:124817-124830.

[16] Caruso A, Chessa S, Escolar S, Barba J, López JC. (2021) Collection of data with drones in precision agriculture: Analytical model and LoRa case study, IEEE Internet of Things Journal, 8(22):16692-16704.

[17] Ruirong D, Jianyu Q. (2021) Design of online traffic monitoring system based on STM32 and LoRa, In 3rd International Conference on Intelligent Control, Measurement and Signal Processing and Intelligent Oil Field, Xi'an, China.

[18] Manuel MP, Faied M, Krishnan M. (2024) A LoRa-based disaster management system for search and rescue mission, IEEE Internet of Things Journal, 11(20):34024-34034.

[19] Djurasevic S, Pešović U, Knežević M. (2022) Evaluation of LoRa wireless technology for military, In 10th International Scientific Conference on Defensive Technologies, Belgrade, Serbia.

[20] Sandoval RM, Garcia-Sanchez AJ, Garcia-Haro J. (2019) Optimizing and updating lora communication parameters: A machine learning approach, IEEE Transactions on Network and Service Management, 16(3):884-895.

https://doi.org/10.31436/iiumej.v26i3.3680

BOUNDARY AWARE SALIENCY-BASED LEVEL SET WITH MOMENTUM CONTRAST METAFORMER FOR SKIN CANCER SEGMENTATION AND CLASSIFICATION

ANJANI GUPTA[1*], SNEHLATA SHEORAN[2], SURABHI RASTOGI[3], ANSHULA GUPTA[4], KAJAL SALUJA[5], SABNAM KUMARI[6]

[1,3,4]Department of Computer Science and Engineering, Vivekananda Institute of Professional Studies-Technical Campus, Delhi, India
[2]Department of Computer Science and Engineering, Chandigarh University, Mohali, Punjab, India
[5]Department of Computer Science and Engineering, Raj Kumar Goel Institute of Technology, Ghaziabad, India
[6]Department of Information Technology, Bhagwan Parshuram Institute of Technology, New Delhi, India

**Corresponding author: anjani.gupta@vips.edu*

(Received: 5 April 2025; Accepted: 1 August 2025; Published online: 9 September 2025)

ABSTRACT: Skin cancer is considered one of the most widespread and life-threatening cancers and remains a challenging task for dermatologists. This challenge arises due to the small boundaries and regions of affected tissue. Existing methods yield poor classification accuracy due to inefficient classifier performance in recognizing complex patterns. Therefore, an effective skin cancer segmentation and classification method called Boundary-Aware Saliency-based Level Set (BASLS) with Momentum Contrast Metaformer (MC-Metaformer) is proposed in this research. BASLS enables improved lesion segmentation by identifying significant structural components along lesion edges. Using residual connections, MC-Metaformer provides a better gradient path, addressing gradient fading issues during deep feature extraction. Preprocessing is performed on the HAM10000, ISIC-2019, and ISIC-2020 datasets to improve and standardize the data. ResNet50 is used to extract relevant features for classification. Experimental results demonstrate that the proposed MC-Metaformer outperforms the Deep Convolutional Neural Network (DCNN), achieving classification accuracies of 99.58%, 99.32%, and 98.62% on the HAM10000, ISIC-2019, and ISIC-2020 datasets, respectively. These results confirm the robustness and efficiency of the model in accurate skin cancer segmentation and classification.

ABSTRAK: Kanser kulit merupakan salah satu jenis kanser paling meluas dan mengancam nyawa serta masih menjadi cabaran utama kepada pakar dermatologi. Cabaran ini berpunca daripada sempadan dan kawasan kecil tisu yang terjejas. Kaedah sedia ada mempunyai ketepatan klasifikasi yang rendah kerana prestasi pengelasan yang kurang berkesan dalam mengenal pasti corak kompleks. Oleh itu, kajian ini mencadangkan kaedah segmentasi dan klasifikasi kanser kulit yang berkesan dikenali sebagai *Boundary-Aware Saliency-based Level Set* (BASLS) bersama *Momentum Contrast Metaformer* (MC-Metaformer). BASLS membolehkan segmentasi lesi yang lebih baik dengan mengenal pasti komponen struktur penting di sepanjang tepi lesi. Melalui penggunaan sambungan residu, MC-Metaformer menyediakan laluan kecerunan yang lebih baik bagi menangani isu kehilangan kecerunan semasa pengekstrakan ciri mendalam. Pra-pemprosesan dilakukan ke atas set data HAM10000, ISIC-2019, dan ISIC-2020 bagi meningkatkan serta menyeragam data. ResNet50 digunakan bagi mengekstrak ciri relevan untuk klasifikasi. Dapatan eksperimen menunjukkan bahawa MC-Metaformer yang dicadangkan mengatasi Rangkaian Neural Konvolusi Mendalam (DCNN) dengan ketepatan klasifikasi sebanyak 99.58%, 99.32%, dan

98.62% masing-masing pada set data HAM10000, ISIC-2019, dan ISIC-2020. Dapatan ini mengesahkan keteguhan dan kecekapan model dalam segmentasi dan klasifikasi kanser kulit secara tepat.

KEYWORDS: *Lesion Edges, Momentum Contrast, Relevant Features, Saliency-based Level Set, Skin Cancer Segmentation and Classification.*

1. INTRODUCTION

The term 'cancer' refers to a broad group of diseases that affect various parts of the human body. Abnormal growth of cancer cells invades neighboring tissues before metastasizing to other organs [1]. Uncontrolled cellular proliferation throughout the body leads to cell death, making cancer the second leading cause of death worldwide. Annual statistics indicate that cancer causes 9.6 million deaths, accounting for one-sixth of all fatalities [2]. Skin cancer is one of the most dangerous types of cancer and frequently affects individuals. It becomes life-threatening when not diagnosed and treated at an early stage [3]. Delayed detection increases mortality rates, as treatment becomes less effective in advanced stages. The two primary categories of skin cancer are melanoma and non-melanoma [4]. Skin pigment cells, known as melanocytes, produce melanomas within the skin tissue, often spreading to other organs. According to global statistics from 2020, non-melanoma skin cancer is the most prevalent subtype, accounting for 6.2% of all annual cancer diagnoses, although it generally results in lower mortality [5][6]. In contrast, the aggressive behavior of melanoma leads to higher fatality rates, despite its lower occurrence compared to non-melanoma [7]. Early detection of melanoma is critical due to its rapid progression. It typically develops on the back, arms, legs, and facial regions [8]. The primary risk factors for developing skin cancer include prolonged exposure to ultraviolet (UV) radiation, the presence of numerous moles, fair skin with sun sensitivity, and a family history of skin cancer [9].

Skin cancer diagnosis is challenging due to poor image quality, combined with factors such as physician experience, visual limitations, and fatigue. These elements contribute to inconsistent diagnostic outcomes among dermatologists [10]. Medical experts have developed computer systems for automatic diagnosis to assist dermatologists in analyzing dermoscopic images [11]. Automated image analysis provides diagnostic support for clinical decision-making; however, healthcare professionals face difficulties distinguishing between different skin cancer types, as their visual presentations often share overlapping characteristics [12]. The diagnostic accuracy of dermatologists in identifying skin cancer ranges from 62% to 80%, depending on their experience level [13]. The dermoscopy technique requires comprehensive training, as new dermatologists need considerable time to achieve precision in their diagnoses [14]. Artificial Intelligence (AI) and Machine Learning (ML) algorithms have driven significant progress in both medical image analysis and disease diagnosis [15]. Enhancements in medical diagnostic systems using Deep Learning (DL) algorithms have led to their widespread adoption across medical research fields [16]. The accuracy of high-performance deep neural networks is improved by expanding Convolutional Neural Networks (CNNs) through increased depth, width, and resolution parameters [17]. However, current skin cancer detection models face challenges in handling small and indistinct lesion boundaries that often blend into surrounding tissue, making accurate segmentation difficult. Traditional models fail to capture fine-grained structural details, especially in noisy or low-contrast images. Additionally, many classifiers underperform due to limited feature extraction and shallow architectures, resulting in poor recognition of complex patterns. This leads to misclassifying visually similar lesions, especially when trained on imbalanced or low-quality datasets, thereby

limiting their generalizability. To address these limitations, this research develops an efficient DL-based approach.

The key findings of this research are listed as follows:

- A Boundary-Aware Saliency-based Level Set (BASLS) with a Momentum Contrast Metaformer (MC-Metaformer) is proposed for effective skin cancer segmentation and classification.
- The BASLS enables improved lesion segmentation by identifying essential structural components along lesion edges. Furthermore, it utilizes adaptive level-set functions to preserve accurate lesion boundaries while maintaining optimal regional uniformity.
- The MC-Metaformer provides an improved gradient path through residual connections during feature extraction. Moreover, the combination of token mixing and depth-wise convolution enhances feature representation, thereby improving classification accuracy.
- In preprocessing, a dull razor removes hair artifacts and improves lesion boundary clarity, while Contrast-Limited Adaptive Histogram Equalization (CLAHE) increases contrast for better lesion visibility. Normalization is applied to scale pixel values, thereby improving model performance.
- ResNet50 extracts deep hierarchical features from skin lesion images using skip connections, which help prevent vanishing gradient issues and enhance model stability. Additionally, it captures complex patterns and textures, thereby enhancing classification performance.

The remainder of this research is organized as follows: Section 2 presents the literature review, Section 3 describes the proposed method, Section 4 provides the experimental results with discussion, and Section 5 concludes the study.

2. LITERATURE REVIEW

The existing research models introduced for skin cancer segmentation and classification are analyzed in this section.

Houssein et al. [18] suggested an efficient multiclass skin cancer classification using a Deep Convolutional Neural Network (DCNN). The DCNN addressed the class imbalance problem and efficiently handled complex visual inputs, rendering it appropriate for image classification. Furthermore, the preprocessing steps were refined to enhance the model's predictive capability with optimal classification accuracy. However, the model's efficiency was limited to smaller datasets, with a deteriorated success rate when tested on larger data quantities.

Ahmad et al. [19] developed a DeepLabv3+ and Vision-based Transformer (ViT) for segmenting and classifying skin lesions. ViT processed information across various image locations in the first layer due to the self-attention mechanism. Furthermore, DeepLabv3+ and ViT were trained from the beginning using selected hyperparameters. However, this method did not account for how different features related to one another, which impaired its accuracy in detecting skin cancers.

Pandey et al. [20] introduced a sparse dictionary-based CNN for skin cancer classification. The Non-Local Means (NLM) denoising and CNN were backed by a sparse dictionary, which utilized a multi-stage approach to improve input data quality, extract meaningful features, and enhance overall classification performance. The combined approach addressed challenges such

as feature extraction, noise robustness, and classification accuracy, making it more suitable for intricate image analysis. However, training this model was difficult due to gradient value decay.

Khan et al. [21] presented an early-stage melanoma diagnosis method from dermoscopic images using a DenseNet and CNN-based approach. The developed model included CNN, DenseNet, batch normalization, max pooling, and a Rectified Linear Unit (ReLU) layer to address the overfitting issue. Moreover, numerous test samples were included, and data augmentation was applied to address class imbalance. The Adam optimizer managed large datasets, effectively handling small gradients. Despite its strengths, this method produced lower-quality results when processing small lesions.

Pandimurugan et al. [22] proposed a DCNN-based model for the early identification and classification of melanoma skin cancer. Segmentation techniques were used to delineate cancer lesions, and DL in computer vision assisted in classifying these lesions. However, automatic detection of melanoma from dermoscopic samples to determine lesion classes was challenging. The model exhibited lower performance due to the images displaying inconsistent shapes and border characteristics.

From the overall analysis, each method proved ineffective in feature selection, as they chose poor-quality features. Furthermore, existing models produced low-quality results when evaluating smaller lesions. This research introduces MC-Metaformer technology to identify different types of skin cancer. ResNet50 captures both low-level texture details and high-level patterns using residual learning, thereby enabling the extraction of discriminative and high-level features. Furthermore, the Boundary-Aware Saliency-based Level Set (BASLS) segmentation enhances boundary localization by preserving complex structural components, particularly in small and indistinct lesions. This segmentation and feature representation improvement enables the model to perform superiorly on small lesion regions, as validated by higher accuracy across the HAM10000, ISIC-2019, and ISIC-2020 datasets.

3. PROPOSED METHODOLOGY

Identifying different types of skin cancer is challenging due to complex patterns and indistinct feature lines. This research introduces a novel DL method for skin cancer segmentation and classification. The proposed model involves five key steps: data collection, preprocessing, segmentation, feature extraction, and classification. The overall procedure for skin cancer segmentation and classification is illustrated in Fig. 1.

https://doi.org/10.31436/iiumej.v26i3.3680

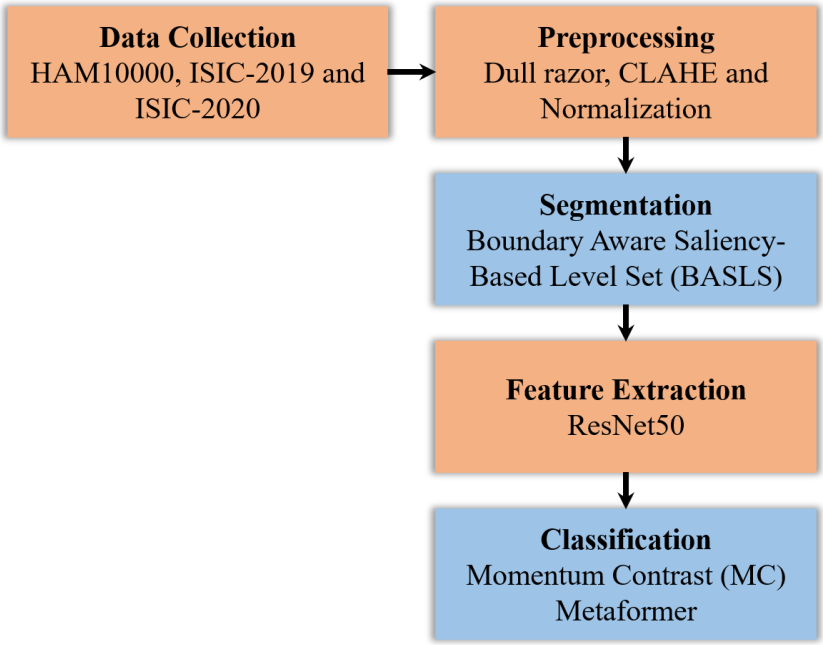


Figure 1. Process for skin cancer segmentation and classification

3.1. Data Collection

This research considers the HAM10000 [23], ISIC-2019 [24], and ISIC-2020 [25] datasets for skin cancer segmentation and classification. The following section provides a detailed description of these datasets.

3.1.1. HAM10000 dataset

The HAM10000 is a publicly available and widely employed dataset for skin cancer classification. It contains 10,105 images divided into seven classes: melanoma (MEL), Melanocytic Nevus (MN), Basal Cell Carcinoma (BCC), Actinic Keratosis (AK), Benign Keratosis Lesions (BKL), Dermatofibroma (DF), and vascular lesions (VL). The class distributions and sample images are listed in Table 1.

Table 1. Class distributions of the HAM10000 dataset

Classes	MEL	MN	BCC	AK	BKL	DF	VL
No. of samples	1113	6705	514	327	1099	115	142

3.1.2. ISIC-2019 dataset

ISIC-2019 is a widely used dataset published by the International Skin Imaging Collaboration (ISIC). It comprises 25,331 images divided into eight classes: Actinic Keratosis (AK), Basal Cell Carcinoma (BCC), Benign Keratosis (BKL), Dermatofibroma (DF), Squamous Cell Carcinoma (SCC), Vascular Lesion (VASC), Melanoma (MEL), and Melanocytic Nevus (NV). The class distribution and number of samples are listed in Table 2.

Table 2. Class distributions of the ISIC-2019 dataset

Classes	AK	BCC	BKL	DF	SCC	VASC	MEL	NV
No. of samples	867	3323	2624	239	628	253	4522	12875

3.1.3. ISIC-2020 dataset

ISIC-2020 is a publicly available and widely used dataset published by ISIC. In this dataset, histopathology is applied to confirm malignant diagnoses, while continuous monitoring and specialist consultation are used to verify benign cases. It includes 33,126 images categorized into seven classes: MEL, Seborrheic Keratosis (SK), Solar Lentigo (SL), Lentigo NOS (LNOS), Lichenoid Keratosis (LK), Nevus, and Unknown. The class distributions and sample images are presented in Table 3.

Table 3. Class distributions of the ISIC-2020 dataset

Classes	MEL	Nevus	LNOS	LK	SL	SK	Unknown
No. of samples	584	5193	44	37	7	135	27126

3.2. Preprocessing

The three datasets are preprocessed using a dull razor, CLAHE, and normalization, as explained in the following subsection.

Dull razor: This method removes hair from skin images to enhance their effectiveness for further analysis. Dull Razor operates according to the following process. 1) A standardized technique identifies dark hair areas across various image backgrounds. 2) The model utilizes bilinear interpolation to replace detected hair by accounting for its thin and elongated characteristics. 3) An adaptive median filter is applied to the replaced pixels, enhancing image clarity.

CLAHE: This method improves local contrast enhancement using the basic Histogram Equalization model. CLAHE does not operate on individual image pixels but instead adjusts contrast within specific regions. Bilinear interpolation handles transitions between these regions and reduces visual distortions. However, CLAHE may introduce unnatural visual effects by boosting contrast, adding unwanted artifacts to the original scenes.

Normalization: Scaling pixels across all areas prevents any feature from becoming more visible than the others [19]. This technique makes training easier by making models less unstable and helping them achieve consistent learning progress. The model becomes more efficient at handling various medical images by reducing its dependence on exact brightness levels.

3.3. Segmentation

The preprocessed images are given as input to BASLS for skin cancer segmentation. The traditional saliency-based segmentation methods segment images to produce maps based on their texture and color. However, the recognition of accurate cancer region boundaries by saliency-based segmentation is affected by low intensity. The level set detects boundaries without color and texture data. However, segmentation using a level set is affected by illumination problems. Since the level set requires data on the cancer forms, and saliency requires boundary indicator functions, they are integrated into this study as BASLS. Furthermore, the traditional level set utilizes an edge detector, but fails to detect accurate boundaries due to lower intensity and illumination issues. Hence, a boundary indicator function is applied to predict accurate skin cancer boundaries. Furthermore, the contour is incorporated as the zero level set of functions. Let φ be the level set determined by the domain Ω, Ω_0 is determined as the zero level set, while Ω_{in} is the region inside Ω_0 and the region outside Ω_0 is defined as Ω_{out}. The energy function $E(\varphi)$ is given by Eq. (1) Where, ε_{img} is the external

https://doi.org/10.31436/iiumej.v26i3.3680

energy determined by the image attribute, ε_{reg} is the internal energy applied as a level set growth constraint, and g_p is an edge detector, which is given in Eq. (2).

$$E(\varphi) = \varepsilon_{img}(\varphi, g_p) + \varepsilon_{reg}(\varphi, g_p) \tag{1}$$

$$g_p = \frac{1}{1+\frac{1}{2}(1-|\nabla I_\sigma|^2\rho^2)(|\nabla I_\sigma|^2\rho^2)} \tag{2}$$

where, I_σ denotes the image smoothed by a Gaussian filter with a standard deviation, and ρ is a threshold boundary function obtained from the standard deviation of the image to regulate the adaptive stopping speed. The boundary threshold function is mathematically expressed in Eq. (3).

$$\rho(I) = \frac{1+\sqrt{S(I_\sigma)}}{3} \tag{3}$$

The derived edge detector, such as the boundary indicator function, is applied in the gradient descent expression, as shown in Eq. (4), comprising three parts. The first part determines the distance-regularized term, which is applied to remove re-initialization. The second part provides long-term driving for the zero-level set model of target boundaries. The final term is an area created through a region-growing matrix and the boundary indicator function. This third part increases the region energy among neighboring targets and enhances the evolution rate.

$$\frac{\partial\varphi}{\partial t} = udiv\left(d_p\left(|\nabla_\varphi|\right)\nabla_\varphi\right) + \lambda\delta_\varepsilon(\varphi)div\left(g_\rho\frac{\nabla_\varphi}{|\nabla_\varphi|}\right) + \left(\alpha g_p + mY\right)\delta_\varepsilon(\varphi) \tag{4}$$

where, u, λ, α and m are weight coefficients which manage every parameter effect, $div(.)$ is a parameter applied to estimate the deviation, ∇ is a gradient of the level set received through the gradient operator, and δ_ε is the Dirac function, as mathematically expressed in Eq. (5).

$$\delta_\varepsilon(\varphi) = \begin{cases} \frac{1}{2\varepsilon}\left(1 + cos\left(\frac{\pi\varphi}{\varepsilon}\right)\right) & |\varphi| \le \varepsilon \\ 0 & |\varphi| > \varepsilon \end{cases} \tag{5}$$

where, δ_ε is obtained from the Heavisine function. The Dirac function is helpful when deployed in conjunction with the integral. The boundary indicator line is integral to an active contour, which is estimated by including the Dirac function, as it achieves boundary adjustment to a zero level set. Subsequently, external energy is attained with the contour acting as a part of the growth stimulus, which is affected by the parameter ε applied to the Dirac function. A large ε enhances the contour-capturing range; however, a higher value reduces contour detection accuracy. Therefore, ε is usually set to 0.5.

3.4. Feature Extraction

The segmented images are passed into the feature extraction process to extract DL features. ResNet50 enhances classification accuracy by capturing significant details from image edges, shapes, textures, and distinctive patterns. Its residual blocks enable the network to learn deeper representations while maintaining stability for better model performance. ResNet50 uses skip connections to address the vanishing gradient problem, allowing the model to learn more effective features. As data progresses through the network, it first detects simple visual elements before identifying more complex object parts in deeper layers.

https://doi.org/10.31436/iiumej.v26i3.3680

3.5. Classification

The extracted features are passed into the skin cancer classification process. The MC algorithm enables unsupervised learning by implementing dynamic dictionary systems for contrastive learning and an adapted Noise-Contrastive Estimation (NCE) loss, which optimizes data representations. The loss function described in subsequent sections serves as the base loss function for this research. The algorithm employs a temperature parameter τ to distinguish a query q from k_+ among $k+1$ options, comparing q to one positive key k_+ and K negative keys. The learning model facilitates information retrieval from both similar and dissimilar image pairs using Eq. (6).

$$L_q = -log \frac{exp(q.k_+/T)}{\sum_{i=0}^{K} exp\ (q.k_i/T)} \tag{6}$$

MC allows for better model expansion by adding fresh mini-batches to the buffer while removing older ones. This technique decouples mini-batch parameters from dictionary size, enabling continuous updates with new mini-batch elements. To stabilize the learning process, the model uses momentum-based updates that gradually modify key parameters θ_k, while multiple query batches are updated with model parameters from the query model θ_q. These optimized parameters follow the gradient descent steps as shown in Eq. (7).

$$\theta_K \leftarrow m\theta_k + (1-m)\theta_{q'} \qquad m \in [0,1) \tag{7}$$

This updated model allows gradual changes, thereby preserving functionality and rendering the model more stable. After the Transformer, the Metaformer offers a universal design with an input embedder with two residual blocks that combine token mixers and a channel MLP. $Input_Embedding$ converts the input feature map I into N tokens, each with a feature dimension of C as shown in Eq. (8).

$$X = Input_Embedding(I) \tag{8}$$

The processor passes the embedded tokens through multiple Metaformer blocks to extract image features, as shown in Fig. 2. In the network design, each Metaformer block consists of two residual modules with normalization. A token mixer supports interactions between tokens in the first residual module. This mixer functions as an identity mapping or a simple MLP, as shown in Eq. (9).

$$Y_1 = Token_Mixer(Norm(X)) + X \tag{9}$$

A normalization operation, known as $Norm$ appears before the residual modules. The second residual module includes an MLP with two layers and an activation function that follows them. Channel MLP improves feature selection by increasing the number of channels from C to rC. The channel dimension is extended from C to rC using text before returning to C. The operation is presented in Eq. (10).

$$Y_2 = \sigma(Norm(Y_1W_1)W_2) + Y_1 \tag{10}$$

where, $\sigma(.)$ is a function that performs the activation process, and W_1 and W_2 are residual modules builds its output from two learned linear transformation layers.

https://doi.org/10.31436/iiumej.v26i3.3680

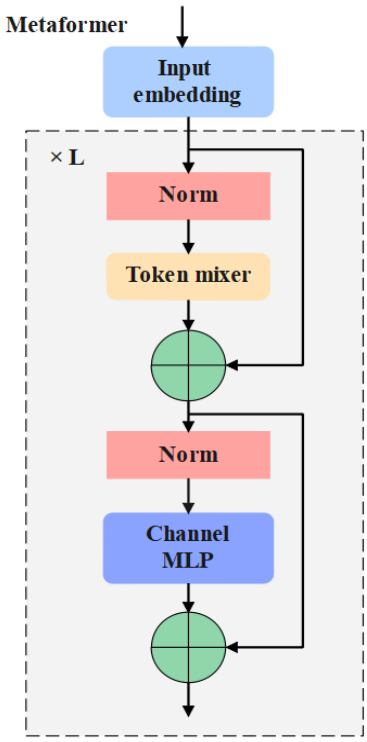


Figure 2. Structure of Metaformer

The proposed novel network design employs a 3×3 depth-wise convolution for token mixing to gather the spatial feature data. This method conserves more computer processing power than self-attention. The operation is defined in Eq. (11).

$$F_S = \delta\left(BN\left(DW(X)\right)\right) \tag{11}$$

where, $\delta(.)$ denotes the LeakyReLU activation element with a depth-wise convolution block. In the Metaformer architecture, the main task of the multilayer perceptron (MLP) is to extract channel-based features, requiring fewer computations than self-attention. The channel MLP component of the token mixer accounts for most of the network's computations when depth-wise convolutions or pooling operate within the token mixer. Experiments with DenseNet and CSPNet demonstrate that adding channel features through concatenation improves network accuracy more than simple element-wise addition. Channel expansion is important in Metaformer to capture more complex and diverse feature information. Metaformer employs a Channel Mixer (CM) module to connect its base blocks to the residual path. The new CM design achieves high overall performance with low computational demands, as shown in Fig. 2.

When the CM module is used, the first linear projection layer transforms the number of channels into $(r-1)C$. The produced channel diversity feature F_C combines the generated channels with the incoming features. The lower-layer feature joins with the spatial result F_S to form a mixed feature O_1 combines the texture details of the upper layers with different types of information. A linear projection layer compresses the features in the channel dimension and selects O_2 based on relevance, as shown in Eqs. (12) - (14).

$$F_C = \delta\left(BN\left(Conv_1(F_S)\right)\right) \tag{12}$$

$$O_1 = Concat\left(F_C, F_S\right) \tag{13}$$

https://doi.org/10.31436/iiumej.v26i3.3680

$$O_2 = \delta\left(BN\left(Conv_2(O_1)\right)\right) \tag{14}$$

where, $Conv_1$ and $Conv_2$ are 1×1 convolutional layers, respectively. CM normalization employs BN as its normalization technique, which runs faster than Layer Normalization (LN) and is a typical CNN component. This model uses a ratio of 2 between the input and output sizes to achieve better results.

4. RESULT ANALYSIS

This section describes the experiments conducted to evaluate the proposed method, including model construction and performance measurement based on specific standards.

4.1. Experimental setup and evaluation metrics

The proposed MC-Metaformer runs on an 8GB RAM and an Intel Core i5 processor that uses the Windows 10 operating system. The developed model measures its performance based on accuracy, F1-score, recall, precision, Dice coefficient, and Jaccard index metrics. Table 4 lists the evaluation criteria and corresponding formulas to determine the model performance. Where, TP, TN, FP and FN denote the True Positive, True Negative, False Positive, and False Negative.

Table 4. Evaluation metrics

Performance metrics	Formula
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
F1-score	$2 \times \frac{precision \times recall}{precision + recall}$
Recall	$\frac{TP}{TP + FN}$
Precision	$\frac{TP}{TP + FP}$
Dice coefficient	$\frac{TP \times 2}{TP \times 2 + FP + FN}$
Jaccard	$\frac{TP}{TP + FP + FN}$

4.2. Segmentation Performance Analysis

This section evaluates different segmentation methods through state-of-the-art techniques to verify their outcomes. The BASLS segmentation model is tested against the GAC, DRLS, and SLS models to determine its effectiveness in detecting skin lesion boundaries. Table 5 presents the segmentation performance of the proposed approach. The proposed BASLS method shows superior segmentation performance compared with other methods, as shown in Table 5. The method achieves 99.58% accurate segmentation of skin lesion areas on HAM10000, producing better results than GAC (93.23%), DRLS (95.37%), and SLS (97.13%). For ISIC-2019 and ISIC-2020, BASLS produces 99.32% accuracy, outperforming GAC, DRLS, and SLS. When segmenting ISIC-2020, BASLS achieves 98.62% accuracy, outperforming all other methods, including GAC (92.65%), DRLS (94.67%), and SLS (96.53%). BASLS generates better lesion segmentation by identifying and retaining important structural components when detecting lesion edges. This approach uses adaptive level-set functions to maintain accurate lesion boundaries with better regional uniformity. The optimized energy function helps BASLS avoid excessive and incomplete segmentations, leading to

https://doi.org/10.31436/iiumej.v26i3.3680

improved scores in the Dice and Jaccard coefficient measurements. Understanding lesion boundaries during processing helps the model detect more lesions in different image sets

Table 5. Segmentation performance for all three datasets

Dataset	Method	Accuracy (%)	Recall (%)	Dice coefficient (%)	Jaccard (%)
HAM10000	GAC	93.23	92.81	88.53	81.63
	DRLS	95.37	94.82	91.36	85.27
	SLS	97.13	96.85	94.24	89.41
	BASLS	99.58	99.19	97.83	95.12
ISIC-2019	GAC	92.51	92.13	87.24	80.42
	DRLS	94.94	94.56	90.72	84.38
	SLS	96.84	96.43	93.53	88.07
	BASLS	99.32	98.87	97.65	94.82
ISIC-2020	GAC	92.65	91.74	86.89	79.92
	DRLS	94.67	94.18	90.23	83.57
	SLS	96.53	96.26	93.08	87.54
	BASLS	98.62	98.12	97.31	94.03

4.3. Feature Extraction Performance Analysis

This section presents the results of various feature extraction techniques compared to current state-of-the-art approaches. It compares the ResNet50's performance against that of the VGG16, DenseNet121, and ResNet18 baseline models, evaluating the efficiency of each structure in utilizing the preprocessed output to discover meaningful features. ResNet50 proves to be an efficient feature extractor, as shown in Table 6. The findings of this research indicate that ResNet50 works best in identifying useful image information. ResNet50 achieves 99.58% compared to 91.45% for VGG16, 94.72% for DenseNet121, and 96.81% for ResNet18 operating on the HAM10000 dataset. Under ISIC-2019 testing, ResNet50 delivered an accuracy of 99.32% above VGG16 at 90.88%, Densenet121 at 93.92%, and ResNet18 at 95.94%. ResNet50 achieved 98.62% accuracy on the ISIC-2020 dataset, surpassing VGG16 (89.77%), DenseNet121 (92.35%), and ResNet18 (94.57%). ResNet50 distinguishes itself from other networks by capturing surface-level visual details and complex lesion structures. Its deep structure contains several convolutional layers, which effectively help learn the visual characteristics at different levels.

Table 6. Feature extraction performance for all three datasets

Dataset	Method	Accuracy (%)	F1-score (%)	Recall (%)	Precision (%)
HAM10000	VGG16	91.45	91.12	90.98	91.07
	DenseNet121	94.72	94.39	94.15	94.22
	ResNet18	96.81	96.52	96.37	96.41
	ResNet50	99.58	99.33	99.19	99.24
ISIC-2019	VGG16	90.88	90.54	90.31	90.47
	DenseNet121	93.92	93.51	93.28	93.42
	ResNet18	95.94	95.68	95.33	95.47
	ResNet50	99.32	98.87	98.53	98.68
ISIC-2020	VGG16	89.77	89.34	89.12	89.24
	DenseNet121	92.35	91.98	91.72	91.86
	ResNet18	94.57	94.31	94.05	94.19
	ResNet50	98.62	98.36	98.12	98.24

4.4. Classification performance analysis

This section describes how the proposed classification method outperforms existing techniques. The MC-Metaformer's performance is evaluated by processing skin cancer types and comparing the results against those of CNN and Metaformer-based approaches for precise classification analysis. Table 7 presents the classification performance of the proposed MC-Metaformer across three datasets. The MC-Metaformer classifier outperforms conventional methods based on the test results. It achieves 99.58% accuracy, surpassing CNN at 93.53%, MC at 95.54%, and Metaformer at 97.56% on the HAM10000 dataset. During the ISIC-2019 evaluation, MC-Metaformer attains an accuracy of 99.32%, exceeding CNN (93.29%), MC (95.31%), and Metaformer (97.35%). For both ISIC-2020 and ISIC-2019 datasets, MC-Metaformer scores over 98% accuracy, outperforming CNN, MC, and Metaformer. The model delivers superior performance by employing enhanced feature extraction and data representation techniques. While processing data, MC-Metaformer effectively detects connections between spatial elements and within channel spaces. It also reduces processing time without losing essential information through its Channel Mixer (CM) setup. The model enhances gradient flow by integrating residual connections, thereby addressing the vanishing gradient problem that occurs during deep feature extraction. The combined use of token mixing and depth-wise convolution architecture further enhances feature understanding, leading to better results across all datasets.

Table 7. Classification performance for all three datasets

Dataset	Method	Accuracy (%)	F1-score (%)	Recall (%)	Precision (%)
HAM10000	CNN	93.53	93.27	93.18	93.22
	MC	95.54	95.28	95.15	95.20
	Metaformer	97.56	97.29	97.16	97.23
	MC-Metaformer	99.58	99.33	99.19	99.24
ISIC-2019	CNN	93.29	92.84	92.52	92.68
	MC	95.31	94.86	94.50	94.67
	Metaformer	97.35	96.88	96.49	96.66
	MC-Metaformer	99.32	98.87	98.53	98.68
ISIC-2020	CNN	92.63	92.38	92.11	92.24
	MC	94.68	94.39	94.14	94.26
	Metaformer	96.66	96.39	96.17	96.28
	MC-Metaformer	98.62	98.36	98.12	98.24

4.5. Comparative analysis

The MC-Metaformer demonstrates superior capabilities against the DCNN [18], ViT [19], Sparse Dictionary-Based CNN [20], Densenet-CNN [21], and DCNN [22]. The hybrid approach is tested against other techniques on HAM10000, ISIC-2019, and ISIC-2020 datasets, and the results are presented in Table 8. When a specific value was unavailable, this research used the symbol NA to indicate it. The results demonstrate that the MC-Metaformer surpasses all state-of-the-art methods across performance metrics. DCNN [18], Sparse Dictionary-Based CNN [20], and DenseNet-CNN [21] deliver lower accuracy compared to the MC-Metaformer, which achieves 99.58% accuracy on the HAM10000 dataset. The MC-Metaformer attains 99.32% accuracy on the ISIC-2019 dataset, outperforming DCNN [18], ViT [19], Sparse Dictionary-Based CNN [20], DenseNet-CNN [21], and DCNN [22]. When processing the ISIC-2020 dataset, MC-Metaformer achieves 98.62% accuracy, surpassing existing models such as ViT and DCNN in labeling precision. This method confirms that MC-Metaformer

produces superior results by effectively capturing more discriminative image features compared to other skin cancer classification approaches.

Table 8. Comparative analysis for proposed MC-Metaformer

Dataset	Method	Accuracy (%)	F1-score (%)	Recall (%)	Precision (%)
HAM10000	DCNN [18]	98.5	98.48	98.51	98.56
	Sparse dictionary-based CNN [20]	85.61	NA	NA	NA
	DenseNet-CNN [21]	93.24	69.68	70.59	68.80
	MC-Metaformer	99.58	99.33	99.19	99.24
ISIC-2019	DCNN [18]	97.11	97.08	97.12	97.09
	ViT [19]	97.73	NA	NA	NA
	Sparse dictionary-based CNN [20]	81.23	NA	NA	NA
	DenseNet-CNN [21]	95.70	80.34	80.93	79.76
	DCNN [22]	88.82	NA	NA	NA
	MC-Metaformer	99.32	98.87	98.53	98.68
ISIC-2020	ViT [19]	96.97	NA	NA	NA
	DCNN [22]	93.45	NA	NA	NA
	MC-Metaformer	98.62	98.36	98.12	98.24

4.6. Discussion

The proposed model achieves strong results and demonstrates clear benefits, as analyzed using HAM10000, ISIC-2019, and ISIC-2020 datasets. The MC-Metaformer model outperforms others in multiclass classification and feature extraction capabilities across all three datasets. The ResNet50 architecture effectively identifies features that yield the highest accuracy in disease classification. On the HAM10000 dataset, ResNet50 achieves 99.58% accuracy, surpassing state-of-the-art methods such as VGG16 (91.45%), DenseNet121 (94.72%), and ResNet18 (96.81%). For both ISIC-2019 and ISIC-2020 datasets, the MC-Metaformer outperforms CNN, Metaformer, and MC models. Metaformer alone also shows better performance than previous approaches, with accuracies of 99.32% on ISIC-2019 and 98.62% on ISIC-2020, outperforming CNN (92.63%), MC (94.68%), and Metaformer (96.66%) on ISIC-2020. The MC-Metaformer's superior performance stems from DL techniques that efficiently extract and classify complex skin lesion features. The classifier detects diverse lesion patterns more effectively by leveraging depth-wise convolution and the Channel Mixer (CM). Furthermore, the proposed method surpasses DCNN, ViT, and Sparse Dictionary-Based CNN in skin cancer detection accuracy. Integrating the BASLS segmentation method enhances lesion boundary detection, further improving medical image classification. Overall, the contrastive learning approach with dynamic dictionaries delivers robust results, establishing the proposed methodology as the leading technique for skin cancer diagnosis.

5. CONCLUSION

The MC-Metaformer significantly enhances the accuracy of skin cancer classification, aiding dermatologists in precise early detection. ResNet50 is employed for feature extraction, effectively capturing small and complex image patterns necessary for multiclass classification. The BASLS segmentation method enhances lesion edge delineation, thereby improving the proposed approach's effectiveness in identifying various skin lesion types. This method addresses two key limitations of previous techniques: inaccurate boundary localization and

poor generalization of deep features. The model introduces a novel framework that achieves superior classification performance by integrating BASLS for precise lesion segmentation with MC-Metaformer for robust representation learning. The MC-Metaformer demonstrates excellent performance and outperforms the existing classification models. The results demonstrate that the proposed method outperforms other methods, including CNN, MC, and Metaformer, with accuracies of 99.58%, 99.32%, and 98.62% on the HAM10000, ISIC-2019, and ISIC-2020 datasets, respectively. The consistent performance of CNN, MC, and Metaformer highlights the contribution of this work in optimizing segmentation boundaries and enhancing multiclass discrimination using contrastive learning and depthwise convolution. The proposed model improves both efficiency and accuracy across datasets. Future research will explore other medical datasets with novel feature selection methods for enhanced performance. Incorporating interpretability tools, the model can be improved to make it accessible and valuable for dermatologists, thereby aiding in early skin cancer detection.

REFERENCES

[1] Alabdulkreem E, Elmannai H, Saad A, Kamil IS, Elaraby A. (2024) Deep Learning-Based Classification of Melanoma and Non-Melanoma Skin Cancer. Traitement du Signal, 41(1): 213-223. https://doi.org/10.18280/ts.410117

[2] Mavaddati S. (2025) Skin cancer classification based on a hybrid deep model and long short-term memory. Biomedical Signal Processing and Control, 100A: 107109. https://doi.org/10.1016/j.bspc.2024.107109

[3] Behera N, Singh AP, Rout JK, Balabantaray BK. (2024) Melanoma skin cancer detection using deep learning-based lesion segmentation. International Journal of Information Technology, 16(6):3729-3744. https://doi.org/10.1007/s41870-024-02004-8

[4] Saha S, Hemal MM, Eidmum MZA, Mridha MF. (2025) Knowledge distillation approach for skin cancer classification on lightweight deep learning model. Healthcare Technology Letters, 12(1):e12120. https://doi.org/10.1049/htl2.12120

[5] Teodoro AAM, Silva DH, Rosa RL, Saadi M, Wuttisittikulkij L, Mumtaz RA, Rodriguez DZ. (2023) A skin cancer classification approach using gan and RoI-based attention mechanism. Journal of Signal Processing Systems, 95(2):211-224. https://doi.org/10.1007/s11265-022-01757-4

[6] Toprak AN, Aruk I. (2024) A Hybrid Convolutional Neural Network Model for the Classification of Multi-Class Skin Cancer. International Journal of Imaging Systems and Technology, 34(5):e23180. https://doi.org/10.1002/ima.23180

[7] Mushtaq S, Singh O. (2024) A deep learning based architecture for multi-class skin cancer classification. Multimedia Tools and Applications, 83(39):87105-87127. https://doi.org/10.1007/s11042-024-19817-1

[8] Sabir R, Mehmood T. (2024) Classification of melanoma skin Cancer based on Image Data Set using different neural networks. Scientific Reports, 14(1):29704. https://doi.org/10.1038/s41598-024-75143-4

[9] Mui-zzud-din, Ahmed KT, Rustam F, Mehmood A, Ashraf I, Choi GS. (2024) Predicting skin cancer melanoma using stacked convolutional neural networks model. Multimedia Tools and Applications, 83(4):9503-9522. https://doi.org/10.1007/s11042-023-15488-6

[10] Alhudhaif A, Almaslukh B, Aseeri AO, Guler O, Polat K. (2023) A novel nonlinear automated multi-class skin lesion detection system using soft-attention based convolutional neural networks. Chaos, Solitons & Fractals, 170:113409. https://doi.org/10.1016/j.chaos.2023.113409

[11] Al Mahmud A, Azam S, Khan IU, Montaha S, Karim A, Haque A, Hasan ZM, Brady M, Biswas R, Jonkman M. (2024) SkinNet-14: a deep learning framework for accurate skin cancer classification using low-resolution dermoscopy images with optimized training time. Neural

Computing and Applications, 36(30):18935-18959. https://doi.org/10.1007/s00521-024-10225-y

[12] Subhashini G, Chandrasekar A. (2024) Hybrid deep learning technique for optimal segmentation and classification of multi-class skin cancer. The Imaging Science Journal, 72(8):1043-1064. https://doi.org/10.1080/13682199.2023.2241794

[13] Jaisakthi SM, Mirunalini P, Aravindan C, Appavu R. (2023) Classification of skin cancer from dermoscopic images using deep neural network architectures. Multimedia Tools and Applications, 82(10):15763-15778. https://doi.org/10.1007/s11042-022-13847-3

[14] Verma N, Ranvijay, Yadav DK. (2024) Hybrid of Deep Feature Extraction and Machine Learning Ensembles for Imbalanced Skin Cancer Datasets. Experimental Dermatology, 33(12):e70020. https://doi.org/10.1111/exd.70020

[15] Ahmed N, Tan X, Ma L. (2023) A new method proposed to Melanoma-skin cancer lesion detection and segmentation based on hybrid convolutional neural network. Multimedia Tools and Applications, 82(8):11873-11896. https://doi.org/10.1007/s11042-022-13618-0

[16] Himel GMS, Islam MM, Al-Aff KA, Karim SI, Sikder MKU. (2024) Skin cancer segmentation and classification using vision transformer for automatic analysis in dermatoscopy-based noninvasive digital system. International Journal of Biomedical Imaging, 3022192. https://doi.org/10.1155/2024/3022192

[17] Khan MA, Alam S, Ahmed W. (2025) Enhanced Skin Cancer Diagnosis via Deep Convolutional Neural Networks with Ensemble Learning. SN Computer Science, 6:124. https://doi.org/10.1007/s42979-024-03581-y

[18] Houssein EH, Abdelkareem DA, Hu G, Hameed MA, Ibrahim IA, Younan M. (2024) An effective multiclass skin cancer classification approach based on deep convolutional neural network. Cluster Computing, 27:12799-12819. https://doi.org/10.1007/s10586-024-04540-1

[19] Ahmad I, Amin J, Lali MI, Abbas F, Sharif MI. (2024) A novel Deeplabv3+ and vision-based transformer model for segmentation and classification of skin lesions. Biomedical Signal Processing and Control, 92:106084. https://doi.org/10.1016/j.bspc.2024.106084

[20] Pandey A, Teja MS, Sahare P, Kamble V, Parate M, Hashmi MF. (2024) Skin cancer classification using non-local means denoising and sparse dictionary learning based CNN. Journal of Electrical Systems and Information Technology, 11:36. https://doi.org/10.1186/s43067-024-00162-0

[21] Khan AR, Mujahid M, Alamri FS, Saba T, Ayesha N. (2025) Early-Stage Melanoma Cancer Diagnosis Framework for Imbalanced Data From Dermoscopic Images. Microscopy Research and Technique, 88(3): 797-809. https://doi.org/10.1002/jemt.24736

[22] Pandimurugan V, Ahmad S, Prabu AV, Rahmani MKI, Abdeljaber HAM, Eswaran M, Nazeer J. (2024) CNN-Based Deep Learning Model for Early Identification and Categorization of Melanoma Skin Cancer Using Medical Imaging. SN Computer Science, 5:911. https://doi.org/10.1007/s42979-024-03270-w

[23] HAM10000 dataset link: https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000 (Accessed on March 2025).

[24] ISIC-2019 dataset link: https://www.kaggle.com/datasets/andrewmvd/isic-2019 (Accessed on March 2025).

[25] ISIC-2020 dataset link: https://challenge2020.isic-archive.com/ (Accessed on March 2025).

https://doi.org/10.31436/iiumej.v26i3.3739

IDENTIFYING TECHNICAL AND VOCATIONAL EDUCATION AND TRAINING (TVET) SENTIMENT FROM SOCIAL MEDIA USING MACHINE LEARNING APPROACH

NUR HAFAZAH BINTI SHARIN[1,2*], MIRA KARTIWI[1]

[1]*Department of Information Systems, Kulliyyah of Information and Communication Technology, International Islamic University Malaysia, 53100 Kuala Lumpur, Malaysia*
[2]*Department of Mathematics, Science and Computer, Politeknik Sultan Salahuddin Abdul Aziz Shah, 40150 Shah Alam, Malaysia*

**Corresponding author: nurhafazah@gmail.com*

(Received: 20 May 2025; Accepted: 2 July 2025; Published online: 9 September 2025)

ABSTRACT: Technical and Vocational Education and Training (TVET) has become a key priority for the Malaysian government to enhance the system, better aligning it with industrial demands and workforce needs. The primary priority is to ensure that students and graduates acquire in-demand skills, thereby increasing their employability and creating more attractive job opportunities. Due to rapid technological advancements, social media has emerged as a powerful platform for public discourse where discussions on TVET programs, policies, and perceptions occur extensively. Among these platforms, Facebook is a widely used space for public interactions through posts and comments. This study employs sentiment analysis to analyse TVET-related discussions on Facebook, categorising sentiment into positive, neutral, and negative polarities. The Term Frequency-Inverse Document Frequency (TF-IDF) method is utilised to extract meaningful insights, and six classifiers, comprised of Support Vector Machine (SVM), Naïve Bayes (NB), Decision Tree (DT), Random Forest (RF), K-Nearest Neighbour (KNN), and Logistic Regression (LR), are applied. Using an 80%-20% training and testing split, results indicate that SVM achieves the highest accuracy performance, with a score of 0.62, outperforming other classifiers. Hence, this study provides valuable insights for policymakers and relevant stakeholders in the TVET ecosystem. By leveraging sentiment analysis and machine learning, decision-makers can better understand public perceptions and develop well-informed strategies to realign and enhance the TVET system.

ABSTRAK: Pendidikan dan Latihan Teknikal dan Vokasional (TVET) menjadi keutamaan kerajaan Malaysia bagi meningkatkan sistem agar lebih selaras dengan permintaan industri dan keperluan tenaga kerja. Keutamaan ini adalah bagi memastikan pelajar dan graduan memperoleh kemahiran yang diperlukan, meningkatkan kebolehpekerjaan serta mewujudkan lebih banyak peluang pekerjaan. Kepesatan kemajuan teknologi menyebabkan media sosial muncul sebagai platfom berpengaruh bagi wacana awam di mana perbincangan mengenai program, dasar, dan persepsi TVET berlangsung secara meluas. Antara platfom tersebut, Facebook menjadi medium terbanyak digunakan bagi interaksi awam melalui hantaran dan komen. Kajian ini menggunakan analisis sentimen bagi menganalisis perbincangan berkaitan TVET di Facebook dengan mengkategorikan sentimen kepada positif, neutral, dan negatif. Kaedah Frekuensi Dokumen Terma *Frequency-Inverse* (TF-IDF) digunakan bagi mengekstrak pandangan bermakna dan seterusnya menerapkan enam pengklasifikasi yang terdiri daripada Mesin Sokongan Vaktor (SVM), *Naïve Bayes* (NB), Pokok Keputusan (DT), Rawak *Forest* (RF), *K-Nearest Neighbour* (KNN), dan Regriasi Logistik (LR). Menggunakan peratusan data pembahagian latihan dan ujian sebanyak 80%-20%, dapatan kajian menunjukkan bahawa SVM mencapai prestasi ketepatan tertinggi dengan skor 0.62, mengatasi pengklasifikasi lain. Oleh itu, kajian ini memberi pandangan berharga kepada

penggubal dasar dan pihak berkepentingan dalam ekosistem TVET. Dengan memanfaatkan analisis sentimen dan pembelajaran mesin, penggubal dasar dapat memperoleh pemahaman mendalam tentang persepsi awam dan membangunkan strategi berinformasi bagi menyelaras dan meningkatkan sistem TVET.

KEYWORDS: *TVET, Sentiment analysis, Machine learning*

1. INTRODUCTION

Technical and Vocational Education and Training (TVET) is a specialised education branch focusing on technology, sciences, and practical skills. It covers various education, training, and skill development across production, services, livelihoods, and occupational sectors. It gives students with the necessary knowledge and tools to pursue careers in various sectors, thereby contributing to business and societal progress [1]. More than just an educational framework, TVET plays a crucial role in building a skilled workforce and boosting national competitiveness. In Malaysia, a well-established TVET ecosystem is a key driver of socio-economic development, addressing persistent challenges like dependency on foreign labor and the mismatch between industry demand and workforce supply. As a result, TVET serves as a strategic pillar in preparing individuals for a dynamic and rapidly evolving job market [2].

With the rise of digital communication, social media has become a powerful platform for TVET promotion, allowing stakeholders, including students and the public, to share experiences, exchange ideas, and engage in discussions [3]. A study by [4] highlighted that social media facilitates effective TVET delivery, enabling instructors to transcend traditional boundaries and provide teaching and learning anytime and anywhere. Additionally, platforms like Facebook and Twitter serve as primary channels for knowledge dissemination, often rivaling traditional media [5].

Given the vast volume of text and opinions generated on social media, sentiment analysis has emerged as a valuable approach to decoding public perceptions of TVET. This method enables the extraction of unstructured data to analyse users' opinions, emotions, attitudes, and perspectives on TVET-related topics [6]. Beyond its analytical function, sentiment analysis is an influential decision-making tool, offering insights into public sentiment regarding services, policies, and programs while identifying the emotional tone embedded in textual data [7], [8].

This study is unique in its contextual emphasis on Malaysian TVET, an under-represented field in sentiment analysis literature. Malaysia's TVET ecosystem faces enduring issues, including an unfavourable public perception of TVET as a second-choice tertiary pathway, fragmented governance across several ministries, and a substantial disparity between graduate competencies and industry demands. Hence, this study leverages a machine learning approach to analyse sentiment from social media platforms, assessing public opinion discourse on TVET. By comparing six classifiers, including Support Vector Machine (SVM), Naïve Bayes (NB), Decision Tree (DT), Random Forest (RF), K-Nearest Neighbours (KNN), and Logistic Regression (LR), this study evaluates their effectiveness in classifying sentiments based on performance metrics including accuracy, recall, precision, and F1-score. Through this approach, this study aims to identify the most reliable classifiers for sentiment classification and contribute valuable insights to policymakers, educators, and stakeholders in the TVET ecosystem by realigning TVET strategies, strengthening institutional credibility, and improving graduate preparedness.

2. RELATED WORKS

Sentiment analysis utilising a machine learning approach has been extensively employed across multiple sectors, including banking, politics, hospitality, entertainment, and education. Many researchers have used this approach to derive insights from textual data and inform decision-making. A sentiment study of consumer feedback on mobile banking applications within the finance sector was conducted by [9] to identify key phrases associated with banking services. Their results demonstrated that SVM surpassed NB in all performance metrics, attaining an accuracy of 97.1%, precision of 97.2%, recall of 97.1%, and F1-score of 97.2%. This underscores SVM's exceptional capacity to analyze client input, making it an invaluable tool for banking service providers to improve the user experience.

A political sentiment study examined the public's perceptions of presidential candidates Joe Biden and Donald Trump during the 2020 United States election campaign [10]. The study harnessed Twitter data for sentiment analysis and employed machine learning classifiers such as Logistic Regression (LR), Random Forest (RF), and K-Nearest Neighbors (KNN). LR exhibited the highest performance of the classifiers evaluated, obtaining an accuracy of 74.3%. The study highlights the efficacy of machine learning in examining political speech and forecasting public sentiment patterns.

Understanding public perceptions of services is crucial for promptly and effectively addressing consumer concerns. A sentiment analysis was performed in [11] by collecting reviews from the TripAdvisor application. Their study concentrated on negative feedback and low-rated reviews, utilizing several machine learning classifiers, including Support Vector Machine (SVM), Naive Bayes, XGBoost, Logistic Regression, and Random Forest. Their findings demonstrated that SVM exhibited remarkable performance, with an accuracy rate of 76%, confirming its effectiveness in evaluating internet reviews. This discovery provides substantial insights for strategic decision-making in the hospitality industry, allowing organisations to tackle client problems proactively.

Machine learning was used in [12] to assess sentiment towards Malaysian places of attraction in the tourism sector. The data was gathered from Twitter and comprised Malay-language text processed with the Malaya corpus. The study evaluated three supervised learning classifiers: Support Vector Machine (SVM), Random Forest (RF), and Naive Bayes (NB). Their findings revealed that SVM achieved the highest accuracy, recall, precision, and F1 score. SVM attained the best accuracy at 92.40%, Naive Bayes at 89.65%, and Random Forest at 88.55%.

Sentiment analysis has also been crucial in comprehending students' viewpoints on online education, especially during the COVID-19 pandemic. [13] examined students' perceptions of online learning through machine learning-based sentiment analysis. Their study showed that the Random Forest (RF) classifier surpassed other classifiers such as Naïve Bayes (NB), Support Vector Machine (SVM), and Decision Tree (DT) regarding accuracy, recall, and precision. These findings underscore the capacity of machine learning to augment educational insights, enabling educators to customize online learning experiences according to students' feedback.

3. METHODOLOGY

This study follows a structured multi-step method, as illustrated in Fig. 1. The process consists of data collection, text preparation and pre-processing, feature extraction, sentiment classification using a machine learning approach, and performance evaluation.

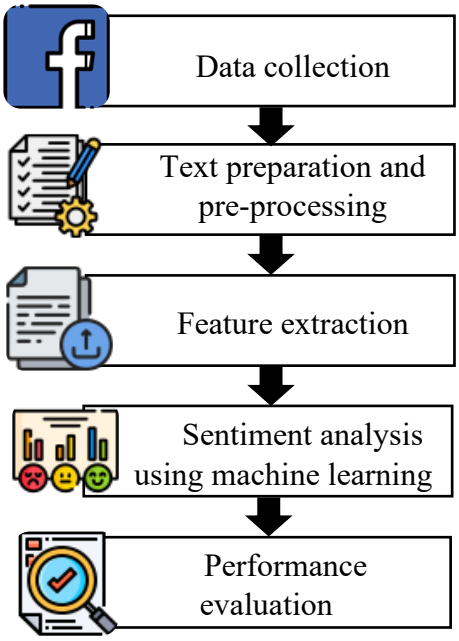


Figure 1. Workflow of sentiment analysis using machine learning

3.1. Data Collection

In this study, the data were collected from publicly accessible Facebook groups and media pages by focusing on contents that engage in active discussions on TVET-related topics. For media pages, only posts and comments specifically addressing TVET issues were included to ensure the relevance of the dataset. The data collection process complied with Facebook's terms of service, and no private user information or personal identifiers were gathered. The data collection process spanned from April 2021 to March 2023 and was facilitated by Facepager and Apify, web scraping tools. Apify, an automated web scraping platform, retrieved and exported unstructured data in .csv format. Meanwhile, Facepager, a tool developed by [14], utilised application programming interface (APIs) and web scraping techniques to retrieve publicly accessible data from YouTube, Facebook, and Twitter.

3.2. Text preparation and pre-processing

Before the analysis process, a data pre-processing stage was carried out to ensure the data quality was suitable for the text mining technique. During data pre-processing, data cleaning and normalization techniques were implemented to improve the quality of sentiment classification. The text preparation process was also carried out at this stage, which involved removing or correcting unnecessary elements such as punctuation, misspellings, abbreviations, and informal language.

The dataset comprised bilingual text written in Malay and English. The Malay data were translated into English using Google Translate to maintain uniformity in sentiment classification. The translated data were further validated using sentiment lexicons to ensure that the sentiment polarity remained consistent and semantically accurate. In addition, the dataset exhibits class imbalance, and no resampling has been implemented to maintain the real sentiment distribution visible in public discussion. This option offers a more reliable classifier performance evaluation in situations similar to real-world sentiment variations. Moreover, the pre-processing phase involved several steps, including lowercasing all text to maintain consistency, removing punctuation and special characters to eliminate noise, and filtering out

stop words to enhance meaningful content extraction. Following this process, 1,304 posts were deemed suitable for analysis.

3.3. Feature Extraction

To convert textual data into a numerical format for machine learning analysis, this study employed the Term Frequency-Inverse Document Frequency (TF-IDF) method. Feature extraction is crucial in transforming raw text into a structured presentation [15]. TF-IDF operates based on two components: Term Frequency (TF), which measures the importance of a word within a document, and Inverse Document Frequency (IDF), which assesses the significance of a word across the entire dataset, thereby reducing the influence of common words [16]. This method was chosen due to its proven effectiveness in improving accuracy metrics and its ability to capture the meaningful importance of words in the text [17].

3.4. Sentiment Analysis using Machine Learning

This study utilised Python-based machine learning libraries to conduct sentiment analysis on the pre-processed text. Six machine learning classifiers were employed, including Support Vector Machine (SVM), Naïve Bayes (NB), Decision Tree (DT), Random Forest (RF), K-Nearest Neighbours (KNN), and Logistic Regression (LR).

3.5. Performance Evaluation

The classification models were evaluated using four key performance metrics as shown in Equations (1) to (4). Accuracy, as expressed in Eq. (1), measures the proportion of correctly predicted sentiments relative to the total number of instances. In addition, precision represented in Eq. (2) is the proportion of correctly identified positive cases among all the instances predicted as positive. Meanwhile, recall evaluates the classifier's ability to correctly identify true positive instances from the total actual positive instances in the dataset as defined in Eq. (3). F1-score provides a harmonic mean between precision and recall, providing a balanced measure of the classifier's overall performance as computed in Eq. (4). By employing these evaluation metrics, the study ensures a comprehensive performance evaluation and facilitates the identification of the most effective machine learning classifier for TVET sentiment analysis where TP denotes the number of correctly classified positive values, TN the correctly classified negative values, FP the incorrectly classified positive values, and FN the incorrectly classified negative values.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{F1} - \text{Score} = \frac{2 \times precision \times recall}{precision+recall} \quad (4)$$

4. RESULT AND ANALYSIS

This study classified 1,304 posts and comments from Facebook using SVM, NB, DT, RF, KNN, and LR classifiers. Referring to the performance of multiple classifiers, SVM attained the highest accuracy of 0.62, closely followed by LR at 0.61. The classifiers of KNN got the lowest accuracy at 0.38, showing their ineffectiveness in categorising TVET sentiment. Among the classifiers, SVM consistently exceeds other classifiers, demonstrating comparable

performance across all sentiment categories. The NB and RF attained an accuracy of 0.55, demonstrating competitive outcomes although exhibiting diverse performance among various sentiment categories. Fig. 2 depicts the accuracy performance of all utilised classifiers, offering a comparative analysis of their effectiveness in sentiment classification.

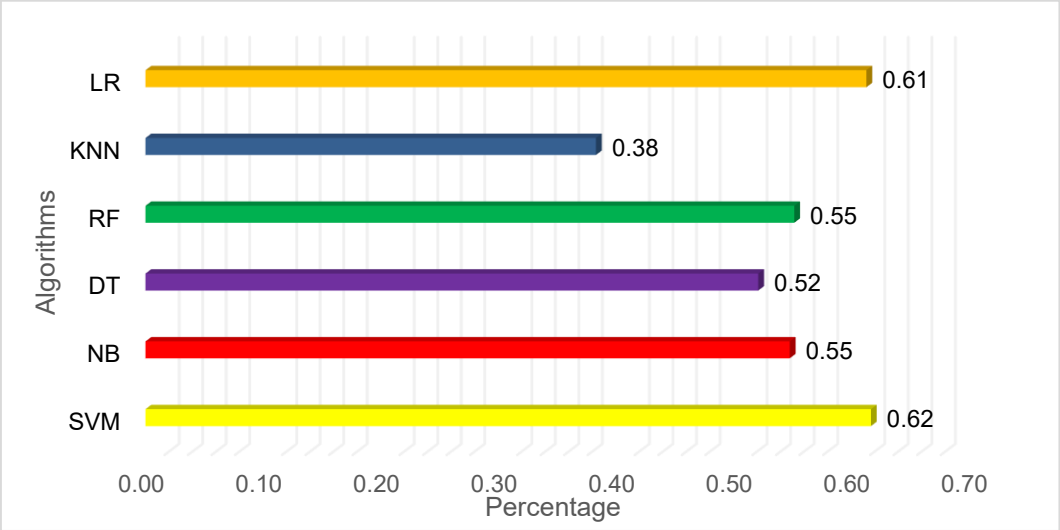


Figure 1. Accuracy performance of machine learning classifiers for TVET sentiment classification

Although SVM demonstrated the highest accuracy, it appears lower compared to prior findings in sentiment analysis studies. Such low accuracy in sentiment analysis may be attributable to many factors, among them the issue of word ambiguity and context dependence. For example, the word 'opportunity' may be positive in the context of education and its value to a person, but negative when describing employment, particularly for specific sectors. Lexicon-based approaches often fail to capture these nuances, leading to misclassification.

Furthermore, the dataset derived from Facebook exhibits many informal terms and languages, bilingual expressions, and sentiment ambiguity. These factors also often impair reliability. The SVM confusion matrix in Table 1 indicates that the most accurate classifications are found along the diagonal, with 36 correctly predicted positive instances, 53 neutral instances, and 71 negative instances. However, misclassification occurred in 16 positive instances, which were erroneously labeled as negative, while the 29 neutral instances were incorrectly predicted as negative. These errors indicate overlapping sentiment expressions, particularly between neutral and other polarities, which are frequently observed in social media discussions.

Table 1. SVM confusion matrix

	Predicted: Positive	**Predicted: Neutral**	**Predicted: Negative**
True: Positive	36	8	16
True: Neutral	15	53	29
True: Negative	14	19	71

A comprehensive analysis of precision, recall, and F1-score for each classifier across positive, neutral, and negative polarities is presented in Table 2. For the positive sentiment, SVM and NB achieved the highest precision scores of 0.74 and 0.70, respectively. However, recall scores for positive polarity remain consistently low across all classifiers, with DT and KNN exhibiting the lowest performance scores at 0.17 and 0.05, respectively. This indicates

that although these classifiers may accurately detect certain instances of positive sentiment, they exhibit poor recall, missing a substantial number of positive examples.

Regarding neutral sentiment, both SVM and LR achieved the highest precision scores of 0.66 and demonstrated comparable performance. In contrast, SVM recorded an F1-score of 0.64, indicating an effective classification of neutral polarity. Meanwhile, the most effective classifier for negative polarity is SVM, with an F1-score of 0.69, followed by NB with 0.63. In contrast, KNN exhibited the lowest performance, achieving an F1-score of 0.04, which highlights its inability to detect negative sentiment effectively.

Overall, SVM is the most effective classifier, recording an overall accuracy of 0.62 and outperforming other classifiers. While LR showed an accuracy of 0.61, it indicated stable performance in sentiment classification. Additionally, NB and RF demonstrated moderate performance, gaining competitive F1-scores, particularly in the detection of negative sentiment. On the other hand, KNN displayed the lowest performance, encountering substantial difficulties in reliably identifying positive and negative sentiment. Based on these findings, selecting classifiers for TVET sentiment analysis must emphasize classifiers that balance precision and recall, as inaccurate classification may result in distorted sentiment representation and misinterpretation of public opinion.

Table 2. Machine learning algorithms evaluation metrics

Algorithm	Acc.	Positive			Neutral			Negative		
		P	R	F	P	R	F	P	R	F
SVM	0.62	0.74	0.47	0.57	0.66	0.62	0.64	0.61	0.78	0.69
Naive Bayes	0.55	0.70	0.27	0.39	0.57	0.44	0.50	0.51	0.81	0.63
Decision Tree	0.52	0.63	0.17	0.25	0.48	0.63	0.55	0.56	0.62	0.59
Random Forest	0.55	0.50	0.18	0.27	0.56	0.55	0.55	0.55	0.77	0.64
K-Nearest Neighbors	0.38	0.60	0.05	0.09	0.38	0.98	0.54	0.67	0.02	0.04
Logistic Regression	0.61	0.60	0.48	0.54	0.66	0.54	0.59	0.59	0.76	0.66

Acc.: accuracy; P: precision; R: recall; F: F1-score

5. CONCLUSION

This study explored sentiment classification using a machine learning approach in Technical and Vocational Education and Training (TVET) discussions on social media. We employed six classifiers, SVM, NB, DT, RF, KNN, and LR, to classify the sentiment into positive, neutral, and negative categories. The Term Frequency-Inverse Document Frequency (TF-IDF) was employed to extract relevant features to enhance the dataset's categorization capabilities.

The results revealed that SVM outperformed other classifiers, achieving the highest accuracy of 0.62 and demonstrating robust performance across all sentiment categories. Logistic regression followed closely with an accuracy of 0.61, while Naive Bayes and Random Forest attained moderate accuracy with 0.55. In contrast, KNN exhibited the lowest performance with an accuracy of 0.38, indicating its inefficacy in handling sentiment classification for TVET discussion. The precision, recall, and F1-score evaluations further emphasised SVM's superiority, particularly in accurately classifying negative and neutral sentiments.

Regarding the insights from the sentiment analysis findings, it is evident that across all classifiers with higher accuracy, positive sentiments towards TVET dominate. This highlights the public's expectations of TVET, particularly its role in helping younger generations access

better education through vocational skills. Further thematic analysis is needed to explore vocational education's perceived benefits and challenges.

Despite its promising results, this study has specific limitations. The dataset is limited to Facebook posts and comments, which may not sufficiently encompass sentiment across other online platforms. Moreover, traditional machine learning models have limitations in addressing the contextual intricacies, sarcasm, and dynamic nature of language in user-generated content. Therefore, future studies should explore the use of a hybrid approach, combining lexicon-based and machine learning methods to enhance sentiment classification performance.

The findings of this study have considerable implications for policymakers, educators, and stakeholders within the TVET ecosystem. Social media sentiment analysis provides strategic value for TVET stakeholders in Malaysia. The collected sentiment trends can assist decision-makers in recognising primary issues, reformulating institutional narratives, optimising curriculum alignment, and ultimately promoting graduate employability. As Malaysia seeks to establish TVET as a leading pathway for industrial preparedness, data-driven sentiment research can effectively align policy with public perception.

REFERENCES

[1] UNESCO-UNEVOC, "What is TVET ?" Accessed: Apr. 04, 2022. [Online]. Available: https://unevoc.unesco.org/home/TVET&context=#ref1

[2] Department of Polytechnic and Community College Education, Hala Tuju Transformasi Politeknik, First edition. Department of Polytechnic and Community College Education, 2024.

[3] D. A. Dutta, "Impact of Digital Social Media on Indian Higher Education: Alternative Approaches of Online Learning during COVID-19 Pandemic Crisis," International Journal of Scientific and Research Publications (IJSRP), vol. 10, no. 05, pp. 604–611, May 2020, doi: 10.29322/ijsrp.10.05.2020.p10169.

[4] F. O. Nwokike, I. C. Ezeabii, and S. N. Oluka, "Social Media: A Platform for Effective Instructional Delivery of Technical and Vocational Education and Training (TVET) Programmes," International Journal of Education, Learning and Development, vol. 9, no. 5, pp. 37–47, 2021, [Online]. Available: https://ssrn.com/abstract=3895718

[5] S. Hashim, A. Masek, N. S. Abdullah, A. N. Paimin, and W. H. N. W. Muda, "Students' intention to share information via social media: A case study of COVID-19 pandemic," Indonesian Journal of Science and Technology, vol. 5, no. 2, pp. 236–245, 2020, doi: 10.17509/ijost.v5i2.24586.

[6] M. Mujahid et al., "Sentiment analysis and topic modeling on tweets about online education during covid-19," Applied Sciences (Switzerland), vol. 11, no. 18, Sep. 2021, doi: 10.3390/app11188438.

[7] F. S. Dolianiti, D. Iakovakis, S. B. Dias, S. Hadjileontiadou, J. A. Diniz, and L. Hadjileontiadis, "Sentiment analysis techniques and applications in education: A survey," Communications in Computer and Information Science, vol. 993, pp. 412–427, 2019, doi: 10.1007/978-3-030-20954-4_31.

[8] R. Hermansyah and R. Sarno, "Sentiment Analysis about Product and Service Evaluation of PT Telekomunikasi Indonesia Tbk from Tweets Using TextBlob, Naive Bayes & K-NN Method," in International Seminar on Application for Technology of Information and Communication (iSemantic), 2020, pp. 511–516.

[9] N. A. Rahman, S. D. Idrus, and N. L. Adam, "Classification of customer feedbacks using sentiment analysis towards mobile banking applications," IAES International Journal of Artificial Intelligence, vol. 11, no. 4, pp. 1579–1587, Dec. 2022, doi: 10.11591/ijai.v11.i4.pp1579-1587.

[10] A. Padwal and R. Koshy, "A Hybrid Approach to Predict Election Candidate Success Using Candidate Speech and Voter Opinion," in Proceedings - International Conference on Communication, Information and Computing Technology, ICCICT 2021, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/ICCICT50803.2021.9510167.

[11] M. S. Shallan, I. F. Moawad, R. El Naggar, and H. Montasser, "Using Machine Learning Techniques to Maximize Profitability in the Hospitality Industry," in 6th International Conference on Computing and Informatics, ICCI 2024, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 182–188. doi: 10.1109/ICCI61671.2024.10485148.

[12] Q. A. B. K. Zaman, W. N. S. B. W. Yusoff, and Q. B. B. A. Shah, "Sentiment Analysis on The Place of Interest in Malaysia," Journal of Advanced Research in Applied Sciences and Engineering Technology, vol. 43, no. 1, pp. 54–65, Jan. 2025, doi: 10.37934/araset.43.1.5465.

[13] S. Madarbakus, A. Gobin, and R. K. Sungkur, "Using Sentiment Analysis and Machine Learning to Collect the Perception of Online Learning," in 2023 Advances in Science and Engineering Technology International Conferences, ASET 2023, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ASET56582.2023.10180757.

[14] J. Jünger and T. Keyling, "Facepager. An application for automated data retrieval on the web. Source code and releases available at https://github.com/strohne/Facepager/."

[15] M. Stanley, K. R. Aiswarya, and G. Deepa, "Sentiment Analysis of Covid Vaccine Tweet with Vader, and Implementation of Different Machine Learning Models," in 2023 14th International Conference on Computing Communication and Networking Technologies, ICCCNT 2023, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICCCNT56998.2023.10308224.

[16] M. Edalati, A. Shariq Imran, Z. Kastrati, and S. M. Daudpota, "The Potential of Machine Learning Algorithms for Sentiment Classification of Students' Feedback on MOOC," 2022.

[17] A. A. Reshi et al., "COVID-19 Vaccination-Related Sentiments Analysis: A Case Study Using Worldwide Twitter Dataset," Healthcare, vol. 10, no. 3, Mar. 2022, doi: 10.3390/healthcare10030411.

https://doi.org/10.31436/iiumej.v26i3.3746

A NOVEL HYBRID MODEL FOR HIGH-ACCURACY MALWARE DETECTION IN THE INTERNET OF MEDICAL THINGS (IOMT) ENVIRONMENT

AMARUDIN DAULAY, KALAMULLAH RAMLI*, DODI SUDIANA, RUKI HARWAHYU, TAUFIK HIDAYAT, NURWAN REZA FACHRURROZI

Department of Electrical Engineering, Universitas Indonesia, Depok, Indonesia

**Corresponding author: kalamullah.ramli@ui.ac.id*

(Received:21 May 2025; Accepted:12 June 2025; Published online: 9 September 2025)

ABSTRACT: The Internet of Medical Things (IoMT) has revolutionized modern healthcare by enabling the collection and analysis of real-time data. However, this interconnected ecosystem also introduces significant security risks, particularly malware attacks that compromise patient safety and data privacy. Traditional security measures are often insufficient because of resource constraints and the real-time operational demands of IoMT devices. This research proposes an optimized hybrid machine learning framework that integrates convolutional neural networks (CNN), long short-term memory (LSTM), random forest (RF), and principal component analysis (PCA) to enhance malware detection in IoMT environments. The proposed method includes an adaptive feature selection mechanism, a resource-efficient architecture, and an ensemble learning model with machine learning capabilities. Validation through experimentation using the CIC-MalMem-2022 dataset, which comprises labeled memory dumps from benign and various malware processes, demonstrated that the proposed framework outperformed current hybrid models while reducing computational costs, achieving a detection accuracy of 99.59%. This study presents a scalable and efficient security solution designed to address the constraints of IoMT devices, addressing critical challenges in healthcare cybersecurity.

ABSTRAK: Internet Benda Medikal (IoMT) telah merevolusikan penjagaan kesihatan moden dengan membolehkan pengumpulan dan analisis data masa nyata. Walau bagaimanapun, ekosistem saling berkaitan ini juga memperkenalkan risiko keselamatan yang ketara, terutamanya serangan perisian hasad yang menjejaskan keselamatan pesakit dan privasi data. Langkah keselamatan tradisional selalunya tidak mencukupi kerana kekangan sumber dan permintaan operasi masa nyata peranti IoMT. Penyelidikan ini mencadangkan rangka kerja pembelajaran mesin hibrid yang dioptimumkan dengan menyepadu Rangkaian Konvolusi Neural (CNN), Memori Jangka Panjang Pendek (LSTM), Rawak Forest (RF) dan Analisis Komponen Prinsipal (PCA) bagi meningkatkan pengesanan perisian Malware dalam persekitaran IoMT. Kaedah yang dicadangkan ini termasuk mekanisme pemilihan ciri penyesuaian, seni bina cekap sumber dan keupayaan pembelajaran mesin bersama model pembelajaran ansembel. Ujian melalui eksperimen menggunakan dataset CIC-MalMem-2022, yang terdiri dari pelupusan memori berlabel daripada proses tidak merbahaya dan pelbagai Malware, menunjukkan bahawa kajian yang dicadangkan mengatasi model Hibrid semasa, juga menurunkan kos pengiraan, mencapai ketepatan pengesanan 99.59%. Kajian ini menyumbang kepada penyelesaian keselamatan berskala dan cekap yang disesuaikan dengan kekurangan peranti IoMT, menangani cabaran kritikal dalam keselamatan siber penjagaan kesihatan.

KEYWORDS: *IoMT, Malware Detection, Machine Learning, Cybersecurity, Healthcare Security*

https://doi.org/10.31436/iiumej.v26i3.3746

1. INTRODUCTION

Digital transformation in the healthcare sector by implementing the Internet of Medical Things (IoMT) has created a global market estimated to reach USD 500 billion by 2030, with an annual growth of 25.9% from 2024 [1]. IoMT enables the connectivity of medical devices, sensors, and applications over the Internet, thereby supporting the real-time collection and analysis of medical data. However, the increased adoption of such technology presents an urgent security challenge. According to the 2024 Cybersecurity and Infrastructure Security Agency (CISA) report, 82% of healthcare facilities have experienced IoMT-related security incidents, reflecting a significant increase and the urgent need for better security systems.

The healthcare sector faces a critical security challenge as malware threatens patient safety and medical data privacy. These sophisticated attacks can severely disrupt essential medical services while exposing sensitive patient information, creating dual risks to healthcare institutions and their patients. The potential consequences include compromised medical device functionality and unauthorized access to confidential medical records. Traditional security approaches are often inadequate for addressing the unique challenges that IoMT devices face, including resource constraints, the need for real-time operations, and the diversity of existing devices and systems [2]. Significant research gaps have been identified, particularly in their ability to detect and adapt to limited operational conditions. Many previous studies have focused on traditional machine learning methods that have not achieved high-accuracy detection of new threats and high false positive rates [3]. This study aimed to develop a malware detection model that is effective in the context of IoMT and efficient in terms of the use of resources, with a detection accuracy target of greater than 99% [4, 5]. According to the CrowdStrike report, interactive intrusion attempts increased by 60% year-over-year, indicating a worrying growth in advanced persistent threats in several industries. While the technology industry has been the primary target, comparable attack vectors are starting to target medical networks, making the healthcare sector, especially IoMT systems, face increasing threats. This highlights the need for stronger cybersecurity solutions tailored to IoMT that can identify and neutralize these dynamic threats [6]. In addition, the latest cybersecurity report revealed alarming trends in attacks targeting healthcare services, where 82% of healthcare facilities reported IoMT-related security incidents by 2024 [7]. This figure highlights the urgent need for robust malware detection systems for IoMT environments. Traditional security approaches often fail to address the unique challenges of IoMT devices, including resource constraints such as limited processing power and memory, real-time operation requirements, device heterogeneity, and continuous data-streaming requirements.

Previous research has explored different approaches to IoMT security [7], including the one-class support vector engine (OCSVM), which achieves 99.3% accuracy [8], the SEResNet50 and BiLSTM, a mindfulness framework, which achieves 97.18% accuracy [5], and immersive features with RF combination, which achieves 99.06% accuracy [9]. However, significant gaps exist, such as the limited ability to detect new malware variants, high computing overhead during training, lack of real-time detection capabilities, and inadequate adaptation to IoMT resource constraints. Therefore, this study seeks to address these gaps by developing a hybrid machine learning architecture that combines a convolutional neural network (CNN), long short-term memory (LSTM), random forest (RF), and principal component analysis (PCA). The goal is to optimize model training efficiency while maintaining high detection accuracy, enabling the detection of previously unknown malware variants, and validating the approach using real-world IoMT malware datasets. In parallel, CNN and LSTM architectures automatically extract features from the input data.

CNN and LSTM methods extract both spatial and temporal, as well as sequential, features. The features extracted from the CNN and LSTM were combined into a single feature vector. The combined feature vector is then dimensionally reduced by PCA to enhance computational efficiency and eliminate redundancy. The dimensionally reduced feature vector is input to a random forest classifier to predict whether the sample is malware. The prediction results obtained using the random forest algorithm are the final outputs of the system that classify the samples as benign or infected with malware. By combining machine learning approaches and dimensionality reduction, the proposed system attempts to achieve high malware detection accuracy while maintaining low computational complexity; thus, it is suitable for implementation on IoMT devices with limited resources. This study makes a significant contribution by developing a hybrid architecture that integrates CNN, LSTM, RF, and PCA for malware detection, as well as adaptive feature extraction methods to detect new malware variants. In addition, the evaluation framework in this study used a real-world IoMT malware dataset to validate the empirical and Friedman test statistics.

The remainder of this paper is organized as follows: Section 2 reviews related research on IoMT security and malware detection. Section 3 describes the methodology and architecture of the proposed system. Section 4 presents the experimental settings and results, discussing the findings and their implications. Finally, Section 5 presents conclusions and suggestions for future research.

2. RELATED WORKS

Recent studies have explored machine learning approaches for malware detection in IoMT environments. Although existing hybrid models have shown promise, they have specific limitations in the context of the IoMT. Despite their incredible accuracy, CNN-LSTM models, as have been reported in some studies [10, 11], demand a large amount of processing power, unsuitable for IoMT devices. In real-world deployments, deep learning techniques in research by [12] demonstrate susceptibility to adversarial attacks. Previous research has shown that PCA and RF combinations are effective at extracting features [13, 14]; however, they lack temporal pattern recognition, crucial for identifying evolving malware. Our proposed architecture addresses these limitations through the integration of resource-optimized CNN-LSTM, designed for IoMT constraints, and PCA-based reduction that maintains detection accuracy while reducing computation. Additionally, RF classification is employed to improve resilience against adversarial attacks common in IoMT environments.

2.1. Network-based detection systems

Almutairi et al. [15] developed a deep-trust network framework with 98.5% accuracy for detecting network intrusions. The proposed method incorporates real-time traffic analysis using principles of federated learning. Similarly, Manimurugan et al. [16] proposed a lightweight authentication protocol designed explicitly for resource-constrained IoMT devices using physical unclonable functions (PUF) for device identification. Xu et al. [17] proposed a feature extraction method that combines static and dynamic analyses, achieving a detection accuracy of 97.8%. Their work emphasizes the importance of selecting relevant features while maintaining computational efficiency. Joshi et al. [18] improved this approach by implementing sensor-based health monitoring with embedded security features. Tahir et al. [19] proposed a hybrid CNN-LSTM architecture that achieved 98.7% accuracy in malware classification. The proposed model demonstrated superior performance in detecting zero-day attacks compared to traditional machine-learning approaches. Building on this, Gull et al. [20]

developed a reversible data-hiding technique incorporating deep learning for IoMT data encryption.

2.2. Ensemble Methods

Egala et al. [21] combined multiple machine learning algorithms with blockchain technology to create a robust framework for device authentication and patient record anonymity. Karmakar et al. [22] enhanced this approach by implementing virtual network functions with integrated security components. Ding et al. [23] developed a cycle-GAN-based encryption network for medical image protection, and Hossen et al. [24] demonstrated the effectiveness of FLE in maintaining data privacy during analysis. Azeem et al. [25] proposed secure message aggregation algorithms for mobile nodes and fog computing environments. Malware in an IoMT environment poses a serious threat that can compromise patient safety and medical data privacy. Table 1 lists the types of malware problems in IoMT.

Table 1. Malware Types Threatening IoMT

Author	Types of Malware	Threat
Rajkumar et al. [26].	Ransomware	Encrypting critical medical data and demanding a ransom to unlock access
Shao Feng et al. [27].	Trojan horse	Impersonation of legitimate software, stealing sensitive medical data, and opening the back door for further attacks.
Abdulraheem et al. [28].	Spy software	Spying on the activities of IoMT devices, illegally collecting patient data, and stealing confidential information
Sri Priyanka et al. [29].	Bot network	Spying on the activities of IoMT devices, illegally collecting patient data, and stealing confidential information

Malware attacks affect healthcare facilities in various ways. In clinical settings, such attacks can disrupt critical medical services, cause data misreading, and damage medical equipment. Impacts on data collection include patient information leakage, medical record manipulation, and loss of vital historical records. The financial consequences include costly recovery systems, operational losses, and regulatory fines. Security challenges arise from resource limitations, including processing power constraints, memory restrictions, and battery life limitations. Legacy systems present additional vulnerabilities due to outdated devices without security updates, obsolete communication protocols, and a lack of modern security features. Integration complexity compounds these issues due to multiple vendor systems, diverse communication protocols, and challenges in interoperability between different platforms.

AI-powered malware has emerged as a sophisticated adversary in the evolving landscape of cyber threats. These advanced threats use machine learning algorithms to continuously adapt their attack patterns, which makes them particularly challenging to detect and counter. The malware can analyze defence mechanisms in real time and modify its behaviour to bypass security controls to create highly targeted and effective infection strategies. In addition to AI-based threats, zero-day exploits continue to pose significant risks. These attacks target previously unknown vulnerabilities in systems and software, leaving organizations particularly vulnerable because no patches or fixes are available during the attack. The success rate of zero-day exploits remains high due to the inherent advantage of attacking undiscovered weaknesses.

https://doi.org/10.31436/iiumej.v26i3.3746

Organizations implement multilayered defense strategies to combat these evolving threats, including technical controls, machine learning detection, and policy controls.

2.3. Technical Controls

IoMT-specific firewalls provide specialized protection for medical devices, network segmentation isolates critical systems to contain potential breaches, and regular security updates close known vulnerabilities and strengthen system defences through machine learning detection. Real-time anomaly detection identifies unusual patterns and possible threats. Behavioral analysis tracks and flags suspicious system activities, and predictive defence capabilities anticipate and prevent emerging attack vectors. Policy control and comprehensive security frameworks establish standardized protection measures, control access management systems, and monitor system usage and incident response planning to ensure quick and effective reactions to security breaches. This combination of technical, analytical, and policy measures creates a robust defence against AI-powered malware and zero-day exploits; however, continuous adaptation remains essential as threats evolve.

Several critical problems can be identified from the table. These include dataset limitations, feature extraction inconsistency, resource optimization, the compromise between accuracy and efficiency, the lack of IoMT specificity, and a critical analysis of IoMT security challenges. A comprehensive analysis identified five fundamental problems that significantly impact the effectiveness of IoMT security research and implementation. These challenges underscore the disparity between theoretical studies and practical applications in healthcare environments. The dataset limitations and the present status of research on IoMT security suffer significant shortcomings regarding dataset quality and relevance. Most previous studies relied on custom datasets that were either general or too narrow to reflect the distinctive features of IoMT situations. This limitation creates a gap between research models and real-world applications. The absence of comprehensive IoMT-specific datasets means that many security solutions are developed and tested using data that do not adequately represent the actual threats and operational patterns found in medical device networks. This misalignment leads to solutions that perform well in laboratory settings but may fail to provide adequate protection in real healthcare environments.

2.4. Feature extraction inconsistency

A critical weakness of current research approaches lies in the inconsistent application of feature extraction methodologies. Many studies either overlook this crucial step or implement it inadequately, resulting in suboptimal model performance. The lack of standardization in feature extraction processes creates challenges when comparing solutions and establishing best practices. This inconsistency affects the reliability of research findings and the practical effectiveness of security solutions in real-world deployments.

2.5. Resource optimization challenges

The complexity of current security models poses significant implementation challenges for IoMT devices. Many proposed solutions require computational resources that exceed those of typical medical devices. The mismatch between model requirements and device limitations creates a substantial barrier to practical implementation, hence the need for optimized models, especially for IoMT devices with limited resources.

2.6. Accuracy Efficiency

A persistent dilemma exists in balancing detection accuracy with operational efficiency. High-accuracy models often require substantial computational resources, making them impractical for real-time threat detection in IoMT environments. The compromise between accuracy and efficiency presents a significant challenge when developing solutions that provide reliable protection and practical functionality. Finding the right balance in IoMT security research remains a considerable challenge. The security models for IoMT currently in use often adopt a comprehensive approach, but fail to address the unique characteristics and requirements of IoMT environments. This lack of specificity results in solutions that may not effectively protect against IoMT-specific threats. The absence of focus on the unique characteristics of IoMT malware and attack patterns leaves healthcare networks vulnerable to specialized threats. There is an urgent need for security solutions that address the unique challenges and threat landscape of medical device networks.

These problems underscore the need for more focused and practical research on IoMT security. Future research efforts must address these fundamental challenges to develop valuable and implementable solutions in healthcare environments. The evolution of IoMT security depends on bridging the gap between theoretical research and practical application while maintaining high protection standards for critical medical infrastructure. This study addresses several key research questions based on the proposed problem formulation. First, we examined how to develop an optimal malware detection model for IoMT contexts that efficiently uses resources while adapting to new threats. Second, we investigated whether our proposed algorithm combination improved detection accuracy, reduced false positives, and maintained real-time performance. Finally, we explored methods to optimize feature extraction, enhancing detection effectiveness, reducing computational overhead, and preserving critical malware characteristics for analysis.

3. RESEARCH METHODS

This section explains the stages of the research, including the introduction, data collection, feature extraction, dimensionality reduction, malware classification, model evaluation, and conclusion and recommendations.

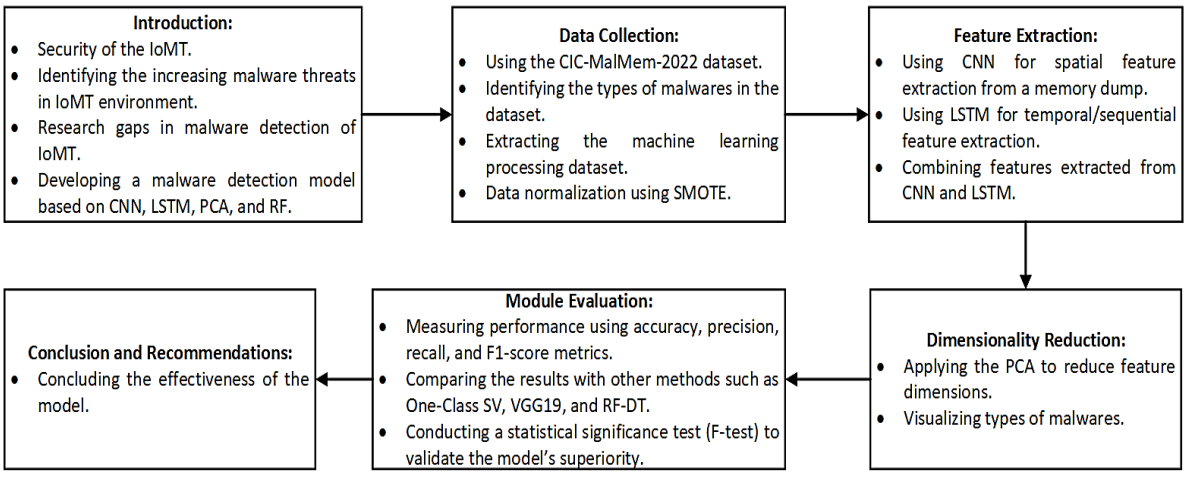


Figure 1. Research Methodology

Figure 1 illustrates the development of a malware detection model for the IoMT. The process began by highlighting the importance of security in IoMT, identifying the growing

malware threats, and pinpointing research gaps related to malware detection. CNN-based models, LSTM, PCA, and RF models were developed to address this issue. The next step was data collection using the CIC-MalMem-2022 dataset, which included malware type identification, dataset extraction for machine learning processing, and data normalization using a synthetic minority over-sampling technique (SMOTE). Following this, feature extraction was performed using a CNN to identify spatial features from the memory dump and an LSTM to capture temporal or sequential features. Then, these two types of features were combined to improve the detection accuracy. The PCA was applied to reduce the dimensionality of the features, which helped visualize various kinds of malware and improved the model's efficiency. After the features were processed and simplified, the data were input into the RF Classifier to determine whether the sample fell into the benign or malware category. The model was evaluated by measuring performance using accuracy, precision, recall, and F1-score metrics. Additionally, the model results were compared with those of other methods to assess the superiority of the proposed approach.

This study used a statistical test approach (t-test) to test the machine learning model. The aim was to evaluate the superiority of the developed model to ensure that the malware detection machine learning model was developed more optimally. The t-test was conducted using Equation 1.

$$T = \frac{mean1 - mean2}{\frac{s(diff)}{\sqrt{n}}} \tag{1}$$

where $mean1$ and $mean2$ are the average values of each sample dataset, $s(diff)$ is the standard deviation of the paired data value differences, n is the sample size (number of paired differences), and $n - 1$ is the degree of freedom. The t-test results conducted in this study also used the Friedman test approach. This test evaluates whether the developed model exhibits significant differences, as expressed in Equation 2.

$$\chi^2 = \left[\frac{12}{N \cdot K \cdot (K+1)}\right] \cdot \left[\sum R_j^2 - 3 \cdot N \cdot (k+1)\right] \tag{2}$$

where N is the number of subjects or blocks, k is the number of treatments or conditions, and R_j is the sum of the rankings for each treatment. j and χ^2_r are test statistics that follow a chi-square distribution with $df = k1$.

3.1. Malware Dataset Preprocessing

Figure 2 shows the preprocessing for the dataset to be managed, and the stages of the malware dataset preprocessing, and displays the steps in processing the CIC-MalMem-2022 dataset [30]. Using this data, attempts are made to categorize malware that is not already categorized, and this study filtered the data to exclude items that were not in use or duplicated from the dataset. After completing this stage, the data was collected in a CSV file for further analysis.

The CIC-MalMem-2022 dataset, developed by the Canadian Institute for Cybersecurity, provides a comprehensive collection of memory dumps from benign and malicious processes. The dataset contains 58,596 records, balanced between benign (29,298) and malicious (29,298) samples. Malicious samples represent three main categories of malware, including Trojan Horse variants: Zeus (195 samples), Emotet (196 samples), Refroso (200 samples), Scar (200 samples), and Reconyc (157 samples); Spyware variants: 180 Solutions (200 samples), CoolWebSearch (200 samples), Gator (200 samples), Transponder (241 samples), TIBS (141 samples); and Ransomware variants: Conti (200 samples), MAZE (195 samples), Pysa (171 samples), Ako (200 samples), and Shade (220 samples). This dataset was particularly valuable

in our research because it represents real-world malware scenarios and includes hidden malware samples; thus, it is ideal for testing detection systems under realistic conditions.

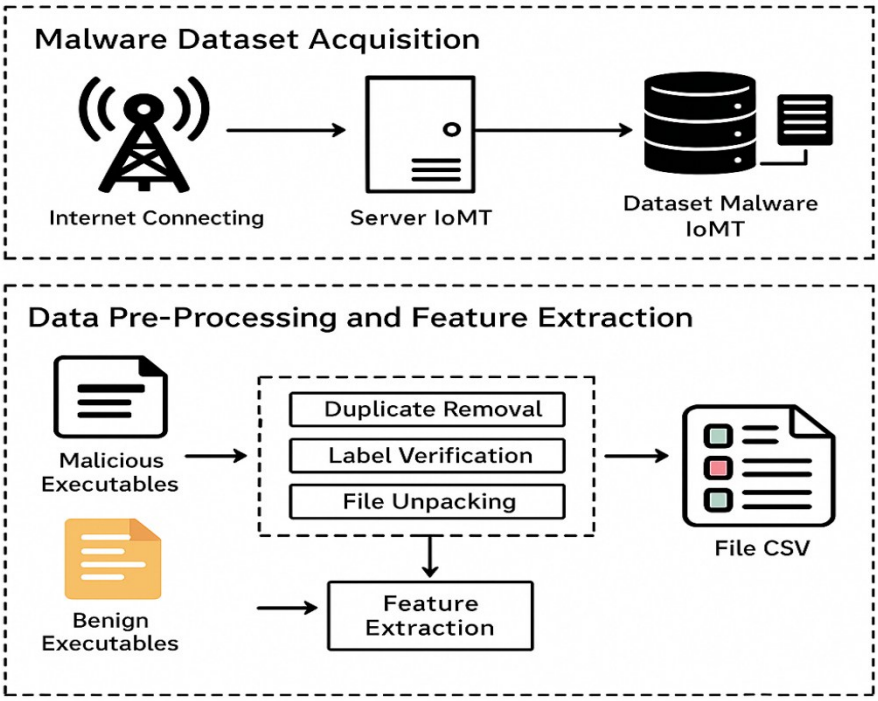


Figure 2. Dataset Preprocessing of Malware

To ensure robust model training and evaluation, we first performed preprocessing, including handling class imbalance using the applied SMOTE for underrepresented malware variants. Then, we used random undersampling for the majority class and maintained a stratified distribution during the train-test split. To limit the simulated resources by reducing the sampling of memory dumps to match the capabilities of devices, artificial network latency and packet loss were introduced to reflect real-world IoMT conditions. Furthermore, data augmentation techniques were applied to account for device heterogeneity. Data cleaning and normalization can remove corrupted memory dumps and incomplete samples, standard feature scaling using min-max normalization, and feature selection based on the resource constraints of IoMT devices.

3.2. Framework of Proposed Method

The proposed hybrid architecture comprises four main components that operate in sequence: a feature extraction layer, a temporal analysis layer, a feature reduction layer, and a classification layer. The CNN architecture was carefully selected after experimental evaluation of multiple configurations, and the three-layer design (32/64/128 filters) proved optimal for IoMT malware detection based on empirical testing, in which single/dual layers showed insufficient feature extraction capability, with an accuracy of 85-90%. Four or more layers increased computational overhead without significant accuracy gains. The selected filter progression (32→64→128) provided the best balance between feature extraction and resource. Each convolution was followed by MaxPooling (2x2) and Dropout (0.25), which were chosen through cross-validation testing that showed a 20% reduction in overfitting compared to alternatives, the temporal Analysis Layer.

https://doi.org/10.31436/iiumej.v26i3.3746

The bidirectional LSTM architecture was selected after comparing various temporal analysis approaches, with a unidirectional LSTM accuracy of 95.8% but lacking the ability to capture reverse sequence patterns. The gated recurrent unit (GRU) showed similar accuracy but a 15% higher computational cost than a bidirectional LSTM with 99.3% accuracy, with only 8% computational overhead. The 128 LSTM units were determined by grid search optimization, which balanced the memory requirements while achieving high detection accuracy. The dropout rate (0.3) was empirically determined to provide optimal regularization without degrading temporal pattern recognition. The feature reduction layer, PCA, reduced dimensionality while preserving information, reducing 256 features to 64 principal components, maintaining 95% of the original variance, and improving computational efficiency. The random forest classifier, with 100 estimators, processed the reduced feature set, outputting binary classification (benign or malicious), and included probability scores for detection confidence.

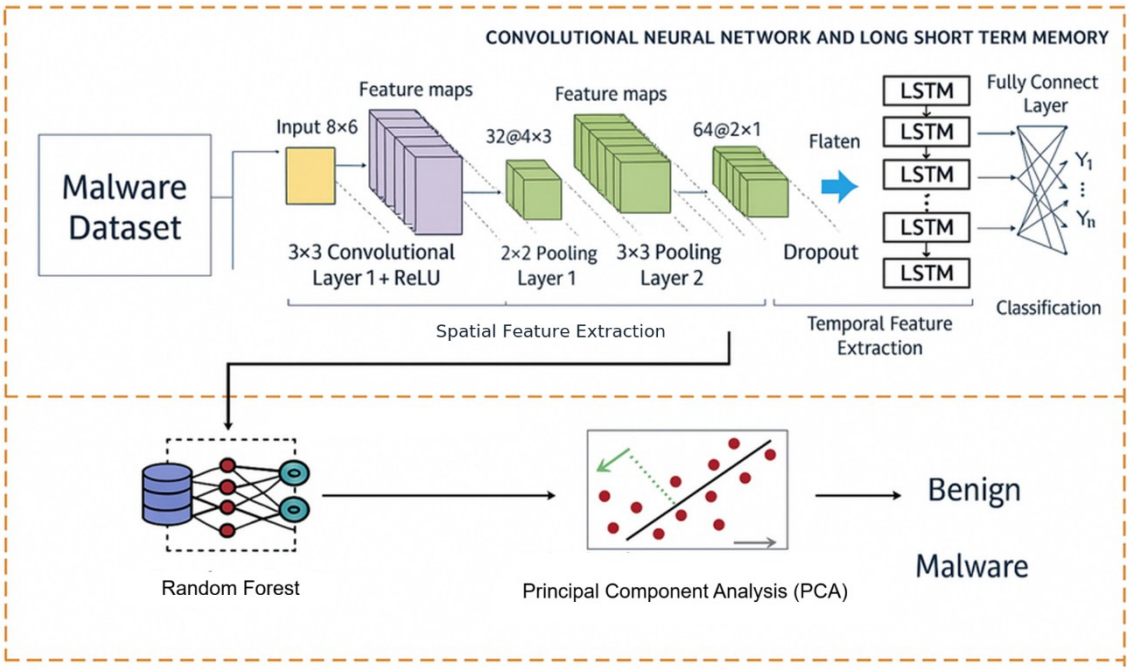


Figure 3. Hybrid Machine Learning Architecture for IoMT Malware Detection

From the malware dataset in the first phase, the CNN and LSTM algorithms were processed to obtain an accuracy of over 10 epochs during testing. After the model became smarter, we attempted to add an RF algorithm and PCA, which were used to determine which malware was not included in the decision-making in the model evaluation. The PCA was used to analyze the spread of malware on IoMT devices; thus, the distribution could be properly visualized. The PCA is a crucial dimensionality reduction technique in the proposed framework.

Implementing PCA in our architecture had several key purposes: dimensionality reduction and computational efficiency. Dimensionality reduction reduced the high-dimensional feature space of memory dump data, transformed correlated variables into uncorrelated principal components, and maintained 95% of the variance while reducing the feature dimensions from 256 to 64. Computational efficiency reduced the training time by approximately 30%, reduced the memory requirements of IoMT devices, and optimized real-time detection capabilities and feature selection.

https://doi.org/10.31436/iiumej.v26i3.3746

4. RESULTS AND DISCUSSION

This research demonstrates a significant breakthrough in IoMT malware detection through an innovative hybrid approach. By combining sophisticated memory dump analysis with dynamic behavioural feature extraction, we achieved an exceptional accuracy rate of 99.59% on the CICAndMal dataset. This remarkable performance can be attributed to three key technological innovations. First, this study implemented a multimodal feature extraction system with comprehensive threat detection capability. Specifically, we considered the following aspects. At its core, the proposed system employed a CNN-based deep feature learning to analyze memory dump patterns, which allowed the detection of subtle malware that might escape traditional analysis methods. LSTM networks processed runtime behaviour, including the detailed monitoring of API call sequences, system call patterns, and network traffic flows. This was further enhanced by tracking resource utilization metrics, which provided insight into suspicious patterns in CPU usage, memory allocation, and I/O operations. Second, we developed a novel fusion architecture that intelligently combined and processed these diverse data streams. The PCA is a foundation for identifying and prioritizing the most discriminative features for each analysis modality. We improved PCA with a custom attention mechanism that dynamically adjusted feature weights based on their demonstrated reliability in threat detection. Architecture culminates in a unique ensemble approach that seamlessly integrates features with traditional machine learning techniques, creating robust and adaptable detection systems. Third, careful optimization addresses the critical challenge of resource constraints in IoMT devices. Our system was implemented with incremental learning capabilities, enabling it to adapt to new malware variants while minimizing the computational overhead of retraining.

The proposed model compression techniques reduced the memory footprint without sacrificing detection accuracy. In addition, selective feature computation was implemented to dynamically adjust the analysis depth according to the available device resources. The effectiveness of the multimodal approach was particularly evident compared with that of single-modality methods. Pure memory analysis alone achieved 99.59% accuracy, whereas isolated behavioural analysis achieved 99.38%. The proposed hybrid approach significantly outperformed both approaches by capturing complementary malware characteristics across different analyses. This superior detection capability was achieved while maintaining real-time performance on resource-constrained IoMT devices, making it a practical solution for healthcare environments.

4.1. Result of Malware Detection Model

The X- and Y-axes represent the two main components of PCA dimensional reduction. The red dot in the sample indicates malware, while the blue dot indicates benign (non-malware) content, and the spread data show a clear clustering pattern. There is an overlap between malware and benign samples around the point of origin. The malware samples exhibited greater variation (spread more widely). Some malware samples were outliers (away from the leading group). Benign samples tended to cluster more tightly in confined areas. These findings suggest potential avenues for further development, including the use of more varied ensemble techniques to enhance accuracy, adaptive thresholding techniques tailored to data clusters, and the selection of nonlinear models to manage overlapping regions better.

Although these improvements were not included in the current model proposal, they are noted as possible directions for further research. Overlapping regions require special handling, malware outliers must be investigated further, and attention should be paid to selecting features to improve segregation. PCA visualization offers crucial insights into the data characteristics

and key areas to consider when developing malware detection models. Although good separation was observed, there were challenges in handling cases in the overlapping and outlier areas (Figure 4).

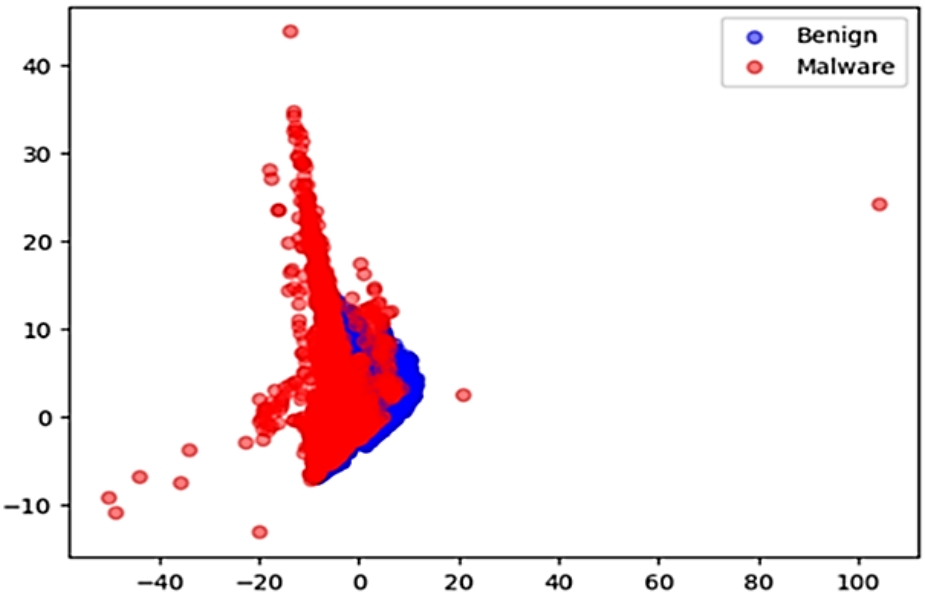


Figure 4. PCA Malware analysis

As shown in Figure 4, a two-sample t-test was performed to ensure statistical significance, ruling out the possibility that the results were due to chance. The results of the comparison between the proposed model and other malware detection methods (one-class SVM and VGG19) demonstrated a $p < 0.05$, indicating a statistically significant difference in accuracy between the proposed model and the different models. Table 3 presents the significance of the test results obtained from the malware detection models. In a comparative analysis, the proposed model outperformed several existing approaches. For example, the one-class SVM achieved an accuracy of 99.30%, while the VGG19 achieved 99.06%. Previous research has shown that hybrid approaches are often more effective than traditional methods. The accuracy of several malware detection techniques used in earlier research is summarized in Table 2 for comparison.

4.2. Comparison of Research on Detection of Malware

This study builds upon earlier research on machine learning-based malware to ensure that our findings can be applied to the developed model, thereby enhancing malware detection on IoMT devices. The experimental results demonstrated that the proposed malware detection model achieved an accuracy of 99.59% on the CICAndMal dataset. These results indicate that the proposed system can detect almost all malware variants tested with a very low error rate. The results of this study have several practical implications for malware detection in the IoMT environment. With high detection accuracy, this system can significantly improve the security of health data collected by IoMT devices, which is critical for protecting patient privacy.

Table 2. Comparison of the Accuracy of Results of Various Methods

Author	Algorithm Machine Learning	Dataset	Accuracy (%)
Alazzam et al. [8]	One-Class SVM	Customize	99.30

Shaukat et al.[9]	VGG19, VGG16, and ResNet5	Malimg	99.06
Shaukat et al. [9]	RF, DT	CICAndMal	99.00
Vinayakumar et al. [31]	NB, DT, CNN	Malimg	98.80
Bansal et al. [32]	VGG19: NB, DT, RF	Caltech-101	93.73
Marastoni et al. [33]	LSTM, CNN	Microsoft	98.50
Cui et al. [34]	CNN	Malimg	94.50
Jian et al. [35]	ResNet50	Microsoft	98.31
Bouchaib et al. [36]	VGG16	Microsoft	98.00
Soni et al. [37]	LSTM, One-class SVM	Customize	97.20
Binbusayyis et al. [38]	Autoencoder, One-class SVM	NSL-KDD	97.11
Kim et al. [39]	Autoencoder	Drebin	97.01
Shaukat et al. [30]	RegNetY320, PCA, one-class SVM	Malimg, Virus Share	99.30
Proposed Model	**CNN, LSTM, RF, and PCA**	**CICAndMal**	99.59

Thus, the proposed method can be implemented in devices with low processing power without sacrificing performance. Security Policy Development: These findings can help healthcare institutions develop better security policies by integrating more effective malware detection solutions. The results demonstrate that the proposed approach effectively detects malware with high accuracy and offers potential for practical application in an IoMT environment. This study significantly contributes to cybersecurity in the healthcare industry using a comprehensive analysis and rigorous significance test.

4.3. The Friedman test

Statistical analysis was conducted to evaluate and compare the performance of four models (RF, CNN, LSTM, and hybrid CNN-LSTM-RF). The goal was to determine whether the observed differences in performance metrics were statistically significant. The findings of the Friedman test are shown in Figure 5. The RF, CNN, and LSTM, as well as the suggested hybrid CNN-LSTM-RF, were the models that were compared using the Friedman test. The results showed a Chi-square value of 14.0400 and a significant p-value of 0.002. This indicates that, according to the assessed metric, such as accuracy, at least one set of models performs differently statistically significantly.

A paired two-sample t-test was performed utilizing the accuracy results of the baseline models and the suggested hybrid model to validate these changes further. The findings indicate that the performance differences are statistically significant, with a p-value below the 0.05 cutoff. This corroborates the finding that the suggested hybrid model performs noticeably better than the baseline models. The t-test and the Friedman test support the proposed approach's improved performance. Using the Nemenyi post-hoc test (Figure 5), further analysis produced pairwise p-values. The Nemenyi post-hoc test, used for pairwise comparisons, revealed a statistically significant performance disparity between the RF and LSTM models, with a p-value of 0.0014 (below the 0.05 cutoff). This indicates that, in the context of the assessed metric, RF performs noticeably better than LSTM. Other pairwise comparisons, including RF vs. CNN ($p = 0.61$), Hybrid ($p = 0.61$) vs. CNN, and LSTM vs. Hybrid ($p = 0.61$), did not, however, show statistically significant differences because their p-values were greater than 0.05. These results suggest that, despite the Friedman test indicating high overall diversity among models, paired evaluations reveal statistically equivalent individual performances. The fact that no single model performed noticeably better than the others suggests that more comparison studies using other measures or datasets are required to determine any clear performance benefits among these modelling techniques.

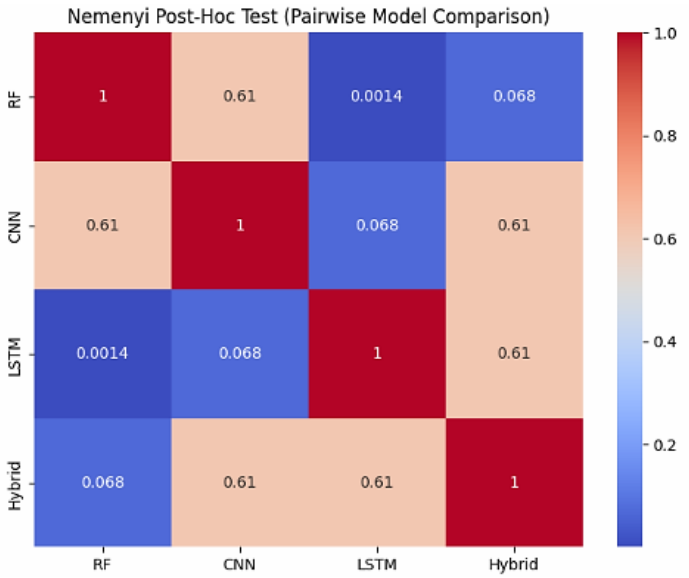


Figure 5. Result of the Friedman Test

5. CONCLUSION

As cyber threats targeting healthcare systems continue to evolve, securing the IoMT has become increasingly critical. This study proposes a hybrid machine learning-based malware detection model that effectively addresses the unique constraints of IoMT environments. The proposed framework employs hybrid methods, including CNN, LSTM, RF, and PCA, significantly enhancing detection accuracy. The empirical results on the CIC-MalMem-2022 dataset demonstrate a superior detection rate of 99.59%, outperforming traditional machine learning approaches. In addition, the resource-aware architecture enables real-time malware detection on constrained IoMT devices. The key contributions of this research include an adaptive feature extraction technique for identifying emerging malware variants, a hybrid architecture that balances detection performance with efficiency, and a validated model suitable for real-world healthcare applications. The comparison of test results obtained using several developed models yields significant findings. Future research should explore federated learning approaches to improve data privacy and investigate adversarial resilience mechanisms to counter sophisticated cyber threats. By advancing IoMT security solutions, this research enhances the protection of critical healthcare infrastructures, ensuring patient data confidentiality and the reliability of medical devices.

ACKNOWLEDGEMENT

This research was supported by the Universitas Indonesia Postgraduate International Indexed Publication Grant 2023-2024 (Number NKB-262/UN2.RST/HKP.05.00/2023).

REFERENCES

[1] G. V. Research, "Internet Of Things In Healthcare Market Size, Share & Trends Analysis Report By Component (Medical Devices, System & Software, Services), By Connectivity Technology, By Application, By End-use, By Region, And Segment Forecasts, 2024 - 2030," 2024. [Online]. Available: https://www.grandviewresearch.com/industry-analysis/internet-of-things-iot-healthcare-market# #.

[2] C. R. Bhukya, P. Thakur, B. R. Mudhivarthi, and G. Singh, "Cybersecurity in Internet of Medical Vehicles: State-of-the-Art Analysis, Research Challenges and Future Perspectives," *Sensors,* vol. 23, no. 19, p. 8107, 2023, doi: 10.3390/s23198107.

[3] V. Ravi, T. D. Pham, and M. Alazab, "Attention-based multidimensional deep learning approach for cross-architecture Iot malware detection and classification in healthcare cyber-physical systems," *IEEE Transactions on Computational Social Systems,* vol. 10, no. 4, pp. 1597-1606, 2022, doi: 10.1109/TCSS.2022.3198123.

[4] L. Dhanya and R. Chitra, "A novel autoencoder-based feature-independent GA optimized XGBoost classifier for Iomt malware detection," *Expert Systems with Applications,* vol. 237, p. 121618, 2024, doi: 10.1016/j.eswa.2023.121618.

[5] R. Abderahman *et al.*, "The internet of things (Iot) in healthcare: Taking stock and moving forward," Elsevier. Internet of Things, 2023.

[6] Crowdstrike, "Global Threat Report 2024," 2024.

[7] M. Humayun, N. Jhanjhi, A. Alsayat, and V. Ponnusamy, "Internet of things and ransomware: Evolution, mitigation and prevention," *Egyptian Informatics Journal,* vol. 22, no. 1, pp. 105-117, 2021, doi: 10.1016/j.eij.2020.05.003.

[8] H. Alazzam, A. Sharieh, and K. E. Sabri, "A lightweight intelligent network intrusion detection system using OCSVM and Pigeon inspired optimizer," *Applied Intelligence,* vol. 52, no. 4, pp. 3527-3544, 2022/03/01 2022, doi: 10.1007/s10489-021-02621-x.

[9] K. Shaukat, S. Luo, and V. Varadharajan, "A novel deep learning-based approach for malware detection," *Engineering Applications of Artificial Intelligence,* vol. 122, p. 106030, 2023, doi: 10.1016/j.engappai.2023.106030.

[10] M. Akhtar and T. Feng, "Detection of Malware by Deep Learning as CNN-LSTM Machine Learning Techniques in Real Time., Symmetry 2022, 14, 2308, ed. Note: MDPI stays neutral about jurisdictional claims in published …, 2022.

[11] T. N. Ghorsad and A. V. Zade, "Hybrid CNN+ LSTM Deep Learning Model for Intrusions Detection Over IoT Environment," *International Journal on Recent and Innovation Trends in Computing and Communication,* vol. 11, pp. 01-11, doi: 10.17762/ijritcc.v11i10s.7588

[12] R. Singhal, M. Soni, S. Bhatt, M. Khorasiya, and D. C. Jinwala, "Enhancing robustness of malware detection model against white box adversarial attacks," in *International Conference on Distributed Computing and Intelligent Technology*, 2023: Springer, pp. 181-196, doi: 10.1007/978-3-031-24848-1_13.

[13] T.-L. Wan *et al.*, "Efficient detection and classification of internet-of-things malware based on byte sequences from executable files," *IEEE Open Journal of the Computer Society,* vol. 1, pp. 262-275, 2020, doi: 10.1109/OJCS.2020.3033974.

[14] E. Safeer, S. Tahir, A. Nawaz, M. Humayun, M. Shaheen, and M. Khan, "Advanced hybrid malware identification framework for the Internet of Medical Things, driven by deep learning," *Security and Privacy,* p. e454, 2024, doi: 10.1002/spy2.454.

[15] S. Almutairi, S. Manimurugan, B.-G. Kim, M. M. Aborokbah, and C. Narmatha, "Breast cancer classification using Deep Q Learning (DQL) and gorilla troops optimization (GTO)," *Applied Soft Computing,* vol. 142, p. 110292, 2023, doi: 10.1016/j.asoc.2023.110292.

[16] S. Manimurugan, S. Al-Mutairi, M. M. Aborokbah, N. Chilamkurti, S. Ganesan, and R. Patan, "Effective attack detection in internet of medical things smart environment using a deep belief neural network," *IEEE Access,* vol. 8, pp. 77396-77404, 2020, doi: 10.1109/ACCESS.2020.2986013.

[17] J. Xu *et al.*, "Healthchain: A blockchain-based privacy preserving scheme for large-scale health data," *IEEE Internet of Things Journal,* vol. 6, no. 5, pp. 8770-8781, 2019, doi: 10.1109/JIOT.2019.2923525.

[18] S. Joshi and S. Joshi, "A sensor-based secured health monitoring and alert technique using Iot," in *2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT)*, 2019: IEEE, pp. 152-156, doi: 10.1109/ICCT46177.2019.8969047.

[19] M. Tahir, M. Sardaraz, S. Muhammad, and M. Saud Khan, "A lightweight authentication and authorization framework for blockchain-enabled Iot network in health-informatics," *Sustainability,* vol. 12, no. 17, p. 6960, 2020, doi: 10.3390/su12176960.

[20] S. Gull, S. A. Parah, and K. Muhammad, "Reversible data hiding exploiting Huffman encoding with dual images for Iomt-based healthcare," *Computer Communications,* vol. 163, pp. 134-149, 2020, doi: 10.1016/j.comcom.2020.08.023.

[21] B. S. Egala, A. K. Pradhan, V. Badarla, and S. P. Mohanty, "Fortified-chain: a blockchain-based framework for security and privacy-assured internet of medical things with effective access control," *IEEE Internet of Things Journal,* vol. 8, no. 14, pp. 11717-11731, 2021, doi: 10.1109/JIOT.2021.3058946.

[22] K. K. Karmakar, V. Varadharajan, U. Tupakula, S. Nepal, and C. Thapa, "Towards a security enhanced virtualized network infrastructure for internet of medical things (Iomt)," in *2020 6th IEEE conference on network softwarization (NetSoft)*, 2020: IEEE, pp. 257-261, doi: 10.1109/NetSoft48620.2020.9165387.

[23] Y. Ding *et al.*, "Deepedn: A deep-learning-based image encryption and decryption network for internet of medical things," *IEEE Internet of Things Journal,* vol. 8, no. 3, pp. 1504-1518, 2020, doi: 10.1109/JIOT.2020.3012452.

[24] M. N. Hossen, V. Panneerselvam, D. Koundal, K. Ahmed, F. M. Bui, and S. M. Ibrahim, "Federated machine learning for detection of skin diseases and enhancement of internet of medical things (IoMT) security," *IEEE Journal of Biomedical and Health Informatics,* vol. 27, no. 2, pp. 835-841, 2022, doi: 10.1109/JBHI.2022.3149288.

[25] M. Azeem *et al.*, "Fog-oriented secure and lightweight data aggregation in iomt," *IEEE Access,* vol. 9, pp. 111072-111082, 2021, doi: 10.1109/ACCESS.2021.3101668.

[26] K. Rajkumar, S. Karthikeyan, V. Kavitha, and U. Hariharan, "Ransomware Attacks in Cyber Space and Mitigation Strategies," in *Cyber Space and Outer Space Security*: River Publishers, 2024, pp. 97-123.

[27] S. Li, H. Zhu, W. Wu, and X. S. Shen, "Backdoor Attacks against Learning-Based Algorithms," ed Springer.

[28] M. Abdulraheem *et al.*, "Artificial Intelligence of Medical Things for Medical Information Systems Privacy and Security," in *Handbook of Security and Privacy of AI-Enabled Healthcare Systems and Internet of Medical Things*: CRC Press, 2024, pp. 63-96.

[29] G. Sripriyanka and A. Mahendran, "Securing Iomt: A Hybrid Model for DDoS Attack Detection and COVID-19 Classification," *IEEE Access,* 2024, doi: 10.1109/ACCESS.2024.3354034.

[30] K. Shaukat, S. Luo, and V. Varadharajan, "A novel machine learning approach for detecting first-time-appeared malware," *Engineering Applications of Artificial Intelligence,* vol. 131, p. 107801, 2024, doi: 10.1016/j.engappai.2023.107801.

[31] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, and S. Venkatraman, "Robust Intelligent Malware Detection Using Deep Learning," *IEEE Access,* vol. 7, pp. 46717-46738, 2019, doi: 10.1109/ACCESS.2019.2906934.

[32] M. Bansal, M. Kumar, M. Sachdeva, and A. Mittal, "Transfer learning for image classification using VGG19: Caltech-101 image data set," *Journal of Ambient Intelligence and Humanised Computing,* vol. 14, no. 4, pp. 3609-3620, 2023/04/01 2023, doi: 10.1007/s12652-021-03488-z.

[33] N. Marastoni, R. Giacobazzi, and M. Dalla Preda, "Data augmentation and transfer learning to classify malware images in a deep learning context," *Journal of Computer Virology and Hacking Techniques,* vol. 17, no. 4, pp. 279-297, 2021, doi: 10.1007/s11416-021-00381-3.

[34] Z. Cui, F. Xue, X. Cai, Y. Cao, G. g. Wang, and J. Chen, "Detection of Malicious Code Variants Based on Deep Learning," *IEEE Transactions on Industrial Informatics,* vol. 14, no. 7, pp. 3187-3196, 2018, doi: 10.1109/TII.2018.2822680.

[35] Y. Jian, H. Kuang, C. Ren, Z. Ma, and H. Wang, "A novel framework for image-based malware detection with a deep neural network," *Computers & Security,* vol. 109, p. 102400, 2021, doi: 10.1016/j.cose.2021.102400.

[36] P. Bouchaib and M. Bouhorma, "TRANSFER LEARNING AND SMOTE ALGORITHM FOR IMAGE-BASED MALWARE CLASSIFICATION," presented at the Proceedings of the 4th International Conference on Networking, Information Systems & Security, KENITRA, AA, Morocco, 2021. [Online]. Available: https://doi.org/10.1145/3454127.3457631.

[37] J. Soni, S. K. Peddoju, N. Prabakar, and H. Upadhyay, "Comparative Analysis of LSTM, One-Class SVM, and PCA to Monitor Real-Time Malware Threats Using System Call Sequences and Virtual Machine Introspection," in *International Conference on Communication, Computing and Electronics Systems*, Singapore, V. Bindhu, J. M. R. S. Tavares, A.-A. A. Boulogeorgos, and C. Vuppalapati, Eds., 2021// 2021: Springer Singapore, pp. 113-127.

[38] A. Binbusayyis and T. Vaiyapuri, "Unsupervised deep learning approach for network intrusion detection combining convolutional autoencoder and one-class SVM," *Applied Intelligence,* vol. 51, no. 10, pp. 7094-7108, 2021, doi: 10.1007/s10489-021-02205-9.

[39] C. Kim, S. Y. Chang, J. Kim, D. Lee, and J. Kim, "Automated, Reliable Zero-Day Malware Detection Based on Autoencoding Architecture," *IEEE Transactions on Network and Service Management,* vol. 20, no. 3, pp. 3900-3914, 2023, doi: 10.1109/TNSM.2023.3251282.

https://doi.org/10.31436/iiumej.v26i3.3788

RELIABILITY TRADE-SPACE EXPLORATION MODELLING FOR SATELLITE ANOMALIES USING EXPONENTIAL DISTRIBUTION

MUHAMMAD IMRAN HARIS AHMAD MAHMOOD[1], NADIRAH ABDUL RAHIM[1]*, ABDUL MUTHOLIB[1,2]

[1]Department of Electrical and Computer Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, 53100, Gombak, Kuala Lumpur, Malaysia
[2]Department of Information Systems, Faculty of Science and Technology, UIN Syarif Hidayatullah Jakarta, Benten 15412, Indonesia

**Corresponding author: nadirahabdulrahim@iium.edu.my*

(Received: 12 June 2025; Accepted: 14 August 2025; Published online: 9 September 2025)

ABSTRACT: The delivery of critical communication, navigation, and earth observation services relies on the reliability of satellite subsystems. However, satellites can be affected by radiation and temperature extremes. Anomalies can cause the system to fail and stop working. This research aims to fix the lack of a high-performance reliability model compared to previous work by integrating trade-space exploration (TSE) techniques and an exponential reliability mathematical model. Data from the Seradata database was analyzed using MATLAB to check for issues with antennas, transponders, amplifiers, and batteries as the prominent components of satellite anomalies. Studies were carried out by building parametric (Weibull, Exponential, and Poisson) and non-parametric (Kaplan-Meier and Monte Carlo Simulation) models, and all were assessed. Results were measured using Root Mean Square Error (RMSE), revealing that the Exponential model performed most accurately compared to other mathematical models. After that, the TSE framework was applied to examine the Design Dependent Parameters (DDPs): reliability, design life, and system performance. The results conclude that the newly developed exponential-based reliability TSE model shows a promising outcome, with RMSE values of 34.2, 10.2, 19.6, and 19.9 compared to Nadirah's model and Shazana's model, which have RMSE values of 16.1, 49.7, 28.2, and 27.6 for antenna, transponder, amplifier, and battery, respectively.

ABSTRAK: Penghantaran Penghantaran perkhidmatan adalah penting dalam komunikasi, navigasi, dan pemerhatian bumi pada kebolehcapaian subsistem satelit. Walau bagaimanapun, satelit boleh terjejas oleh radiasi dan perubahan suhu ekstrem. Anomali boleh menyebabkan sistem gagal dan berhenti berfungsi. Penyelidikan ini bertujuan bagi memperbaiki kekurangan model kebolehcapaian berprestasi tinggi daripada kerja terdahulu dengan mengintegrasi teknik penerokaan ruang dagangan (TSE) dan model matematik kebolehcapaian Eksponen. Data daripada pangkalan data Seradata dianalisa menggunakan MATLAB bagi memeriksa isu antena, transponder, penguat, dan bateri sebagai komponen utama anomali satelit. Kajian dijalankan dengan membina model parametrik (Weibull, Eksponen, dan Poisson) dan non-parametrik (Kaplan-Meier dan Simulasi Monte Carlo). Dapatan kajian diukur menggunakan Ralat Kuadrat Purata Akar (RMSE), menunjukkan bahawa model Eksponen berfungsi paling tepat berbanding model matematik lain. Kemudian, rangka kerja TSE digunakan bagi mengkaji Parameter Bergantung Reka Bentuk (DDPs), yang merangkumi kebolehpercayaan, hayat reka bentuk, dan prestasi sistem. Dapatan kajian menunjukkan bahawa kebolehpercayaan model TSE berasaskan eksponensial yang baru dibangunkan menunjukkan hasil yang memberangsangkan di mana nilai RMSE adalah 34.2, 10.2, 19.6, dan 19.9

https://doi.org/10.31436/iiumej.v26i3.3788

berbanding dengan model Nadirah dan model Shazana adalah 16.1, 49.7, 28.2, dan 27.6 untuk antena, transponder, penguat, dan bateri.

KEYWORDS: *Reliability, Trade-Space Exploration, Exponential, DDPs, RMSE, Satellite Anomalies*

1. INTRODUCTION

Satellites are vital in communication, navigation, remote sensing, and data transmission. Transmission is stabilized in these systems by antennas, transponders, amplifiers, and batteries [1]. Nevertheless, satellites located in space constantly face numerous environmental challenges, including radiation, temperature fluctuations, and micrometeoroid impacts [2]. As a result of these factors, there are often issues in the data and signal subsystems, leading to miscommunications, data failure, or shutting down the entire mission [2].

Many mission failures, decreases in system performance, and higher costs can be traced back to satellite anomalies. Despite the significant effects these failures often have, past studies have not thoroughly examined how reliable satellites are using organized mathematical models. A reliable system design is needed to prevent these risks, but it frequently requires sacrifices on cost, weight, and power use [3]. As a result, designers need to use a structured framework to handle these factors in the design process.

To address these issues, this study investigates satellite subsystem anomalies using standard reliability models within a trade-space exploration (TSE) framework. Therefore, reliability analyses using parametric models, such as the Weibull, Exponential, and Poisson distributions, and non-parametric models, such as the Kaplan-Meier and Monte Carlo simulation, are explored in [4,5]. All the proposed models are measured to identify the best and most helpful way to assess satellite reliability. By integrating TSE and reliability modeling, this study is designed to enhance the design choices made in future satellite systems. Building upon prior reliability analysis work by Imran [5], this study entails the satellite reliability analysis using mathematical prediction models for the case study of satellite anomalies.

2. LITERATURE REVIEW

Satellite communication subsystems help to achieve mission success because they play a vital role in data transmission, telemetry, tracking, and control. Devices such as antennas, transponders, amplifiers, and batteries must often contend with solar radiation, significant temperature fluctuations, and tiny particles in space [1]. Environmental factors are known to create satellite anomalies, which can lower how well a mission performs or bring it to an end. Several studies highlight the relationship between anomalies and the sun, mainly when the Earth's magnetic field is disturbed [3,6].

Designers measure a satellite's dependability by determining its likelihood of completing tasks without error over a specified period. An analysis of actual systems shows that initial failures can happen and that failure rates escalate with time [5]. Therefore, an advanced statistical model is needed to represent these changes [4,7]. The actual reliability formula in this research is as in Eq. (1)[4].

$$R_t = 1 - \frac{[x_t - y_t]}{x_t} \quad (1)$$

where $x_t$is design life in years and y_t is the age since launch.

https://doi.org/10.31436/iiumej.v26i3.3788

Techniques in reliability modeling can be classified as parametric and non-parametric. These models work best when a known failure distribution is applied to predict reliability over time accurately [5]. Using Weibull models, the hazard rate may improve, worsen, or remain constant over time. A system's failures may be modelled with an Exponential or Poisson model, depending on whether they are continuous or if there are intervals [5]. For these methods to work properly, very accurate assumptions are needed, as they depend strongly on the amount and quality of available data [4]. With non-parametric Kaplan-Meier and Monte Carlo methods, it is possible to have great flexibility in modeling reliability when the distribution is unclear or the data is censored [5]. When the Kaplan-Meier estimator is used, survival probabilities can be calculated without making specific distribution assumptions, which can be suitable when the reliability data is incomplete [5]. Another approach is the Monte Carlo simulation, which generates random values to determine the probability of reliability outcomes [5].

2.1. Mathematical Prediction Models: Parametric Reliability

Evaluating performance and possible failures in a system relies on reliability prediction. Weibull and Exponential distributions are curved and fitted for reliability prediction, but the Poisson distribution uses a discrete approach [8-10]. With the aid of these models, engineers can learn about a structure's reliability, design lifetime, and performance [11].

Meanwhile, the Weibull distributions are used to model satellite reliability because they can fit an increasing, unchanging, or decreasing trend [12,13]. Reliability predictions are estimated for a Weibull curve fit on the reliability data [12]. Weibull curve fitting is described in Eq. (2) [5].

$$f(x) = abx^{(b-1)}e^{-ax^b} \tag{2}$$

where x represents the time parameter, a the scale parameter, and b the shape parameter.

On the other hand, the exponential distribution is appropriate when failures are not linked by time [10]. The Exponential curve fitting is presented in Eq. (3) [5].

$$f(x) = ae^{-bx} \tag{3}$$

where a denotes the initial reliability and b the rate of decay.

Discrete failures are modelled using the Poisson distribution for each defined time interval [9]. Reliability data is analyzed using the Poisson discrete probability approach to make reliability predictions. Poisson discrete probability can be seen in Eq. (4) [5].

$$f(x) = \frac{e^{-a}a^x}{x!} \tag{4}$$

where a denotes the rate parameter and x the number of failures.

2.2. Mathematical Prediction Models: Nonparametric Reliability

When the failure distribution type is unknown, non-parametric reliability mathematical prediction models are used instead [14]. Statistics is used to estimate the reliability of the information provided by these methods. Kaplan-Meier and Monte Carlo simulations are used to estimate reliability. At the same time, Monte Carlo Simulation uses probability distributions to make predictions about reliability [8].

The failure probability data is used with the Kaplan-Meier method to get the survival function. Eq. (5) [14] shows the technique for calculating failure probability.

https://doi.org/10.31436/iiumej.v26i3.3788

$$R(t) = 1 - F(t) \tag{5}$$

The Kaplan-Meier Estimator, defined in Eq. (6) [14], is useful for assessing censored data when the failure time is unknown [15].

$$\hat{S}(t) = \prod_{i:t_i \le t}(1 - \frac{d_i}{n_i}) \tag{6}$$

Where $\hat{S}(t)$ is the estimated survival probability at time t, t_i is the observed failure times, d_i is the number of failures at time t_i, and n_i is the number of units at risk just before time t_i.

Meanwhile, Monte Carlo simulation utilizes various conditions to identify possible outcomes [16,17]. It is defined in Eq. (7) [16,17].

$$R(t) = \frac{1}{N}\sum_{i=1}^{N} I(T_i > t) \tag{7}$$

where $R(t)$ is the estimated reliability at time t, N is the total number of simulated trials, and $I(T_i > t)$ is the indicator function equal to 1 if the system survives beyond t and 0 otherwise.

2.3. Trade-Space Exploration

Trade-space exploration (TSE) analyzes various ways to design a system to find the best solution for several performance measures [11,18]. TSE allows designers to examine how reliability, cost, and performance are affected by different constraints. Optimizations in radar, power systems, and structures have been performed in satellite design using TSE. Nevertheless, few studies have managed to link fault tree analysis with standard reliability models, especially on real satellite anomaly data [19].

Researchers have proposed new statistical models and mixed approaches; however, these solutions often have shortcomings due to small datasets, insufficient practical validation, or overly restrictive assumptions [2,5,20]. It indicates that more work is needed to use thorough data-supporting reliability approaches in formal TSE [19]. By filling this gap, users can understand component risks better and build more durable satellite networks [18,19]. In this study, an exponential-based reliability TSE model for a case study of satellite anomalies is chosen to be developed based on the lowest root mean square error (RMSE) values obtained from the four most failed components, which are antenna, transponder, amplifier, and battery [5].

3. RESEARCH METHODOLOGY

The methodology for this study is constructed to analyze satellite reliability by identifying anomalies in satellite communication subsystems and investigating the balance between reliability, cost, and performance, ultimately leading to the development of a traditional mathematical reliability model using trade-space exploration (TSE). Previous research focuses on TSE, and traditional mathematical reliability models are reviewed to identify their applicability and research gaps.

Fig. 1 depicts the research methodology flowchart comprising the steps to perform this study. The first step is to arrange the satellite reliability data based on the satellite anomalies obtained from Seradata [21]. Then, the most prominent anomalies are discovered from the database. The discovered anomalies are divided into four primary communication subsystem failures: antenna, transponder, amplifier, and battery [18]. Approximately 87 satellite reliability data points from various satellites and orbits worldwide, sourced from the Seradata database, are analyzed using MATLAB [5, 21]. After obtaining the satellite reliability data for the four mentioned components, the Design Dependent Parameters (DDPs) identified are reliability,

https://doi.org/10.31436/iiumej.v26i3.3788

design life, and performance. Then, the best mathematical model, which happened to be Exponential, was found to have the best performance in addressing reliability using root mean square error (RMSE) [5]. Then, the model was refined to suit the DDPs. Lastly, this paper applies the TSE method to this mathematical model. A new exponential-based reliability TSE model is developed. This new TSE model must be validated against other existing models, and the RMSE values must be computed.

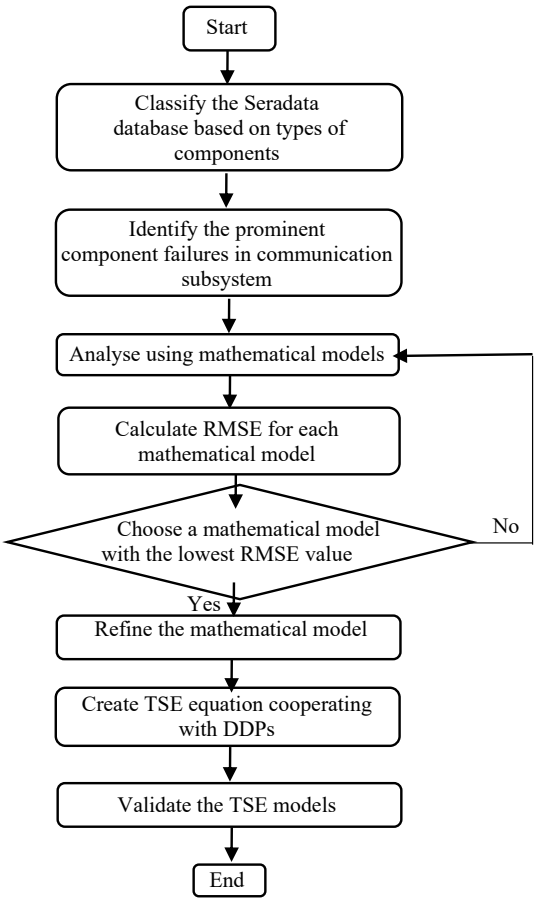


Figure 1. Research methodology flowchart

4. RESULTS AND DISCUSSION

The graphical representations for Weibull, Exponential, Poisson, Kaplan-Meier, and Monte Carlo distributions can be found in [5]. This paper focuses on the development of the exponential-based reliability TSE model for a case study of satellite anomalies. The TSE framework is created to study the correlation between reliability, design life, and performance over time. It makes it possible to match various subsystems and decide on the best choice by assessing their reliability over the system's life.

This method selects a mathematical model that accurately describes the wear and tear of components. At the first phase of analysis, various models, such as Weibull, Poisson, and Exponential, as well as Kaplan-Meier and Monte Carlo functions, were compared with the reliability data gathered from observing component performance [5]. By examining the performance using Root Mean Square Error (RMSE), it was found that the Exponential model best fit the data [5]. The RMSE values obtained from the Exponential model were the lowest, indicating that it fits the actual reliability data the best. The results can be found in Imran's work [5]. According to reliability literature, the exponential function is commonly applied to

systems where the hazard rate remains relatively constant [2]. As a result, the exponential reliability function was adopted and adapted for this study. The classical exponential reliability equation is defined as in Eq. (8) [5].

$$R(t) = e^{-\lambda t} \tag{8}$$

where $R(t)$ denotes the reliability at time t and λ denotes the failure rate of the component.

However, given the high failure rate values in the collected data, the resulting reliability curves turned out to be too steep when this formula was applied. Therefore, to obtain more realistic reliability trends for the entire operational time, the scaling coefficient α was added to the model. The earlier expression can be modified as in Eq. (9).

$$R(t) = e^{-\alpha\lambda t} \tag{9}$$

where $R(t)$ denotes the reliability at time t, λ is the failure rate of the component, and α is the scaling coefficient.

The α parameter helps control the failure rates' role over time. A value of α = 0.001 was chosen in this study to prevent the effects of high failure rates and to demonstrate a more likely rate of reliability decline. If this adjustment were not added, the exponential curve would fail to capture reliability and show irrelevant component performance. In addition, the model was improved by adding a parameter that measured its predictive accuracy. In particular, (1−RMSE) was used as a multiplier in the reliability function to form the final exponential-based reliability TSE developed model as in Eq. (10).

$$TSE_{R(t)} = e^{-\alpha\lambda t}(1 - \text{RMSE}) \tag{10}$$

RMSE indicates the difference between the predicted and actual reliability values to evaluate the model's accuracy. The higher the RMSE, the lower the performance. Because (1−RMSE) is included, models with significant RMSE errors receive a lower reliability prediction than those with lower RMSE errors. Making this modification guarantees that the model's accuracy affects the reliability of the projections. Thus, in this situation, the lower the RMSE, the more reliable the prediction, upholding the goal of picking models supported by practical results and theory. This is the RMSE value obtained from previous research, where the values are 0.22 for antenna, 0.27 for transponder, and 0.30 for both amplifier and battery [5]. For this analysis, the TSE approach relies on three DDPs: design life in years, reliability, as the result of the model function, and RMSE, which represents the model's performance. The strategy developed by the TSE framework makes it easier for engineers to evaluate subsystem components and decide how to design the system. Combining theory, measurement, and priority, the TSE method makes it possible to understand and boost the reliability of satellite parts while choosing and updating the design for engineering needs [11].

The TSE equation aims to evaluate the reliable performance of each subsystem component in the long term using a set approach. Using the same exponential decay model and parameter values for every subsystem allows for a fair and even comparison of systems. The space created by the trade-offs between DDPs demonstrates the reliability of each component and identifies situations where it might be sensible to have higher redundancy or tougher components [4]. Using this method, a better design for satellite subsystems can be chosen with more reliable information. Tables 2 and 3 compare Nadirah's and Shazana's reliability models against the new exponential-based reliability TSE model [2, 20]. The new proposed exponential-based reliability TSE model and two existing reliability models by Nadirah and Shazana are compared in Table 1 for validation purposes. The processes used to determine reliability are assessed over a ten-year design life for the antenna, transponder, amplifier, and battery. The

https://doi.org/10.31436/iiumej.v26i3.3788

findings make it clear that the new TSE model tends to estimate declining reliability over the years. The model now provides more reliable and component-focused estimates of failure, demonstrating greater accuracy in how materials or components wear out over time. For instance, the reliability value in the new model for Antenna is declining steadily from 0.72 in year one to 0.37 in year 10, similar to the other three components. However, the reliability value in Nadirah's model remains high, beginning with 0.94 in year one but still retaining 0.64 in year ten, suggesting slower degradation. Meanwhile, Shazana's model demonstrates a more aggressive decline from 1.01 in year one to 0.34 in year ten, indicating a degradation over time.

Table 1. Validation of Exponential-based reliability TSE model against two existing reliability models [2,20]

Design life	Previous study		New			
	Nadirah	Shazana	Antenna	Transponder	Amplifier	Battery
1	0.93	1.01	0.72	0.77	0.75	0.75
2	0.9	0.89	0.67	0.76	0.72	0.72
3	0.87	0.79	0.62	0.75	0.69	0.69
4	0.84	0.7	0.58	0.74	0.67	0.67
5	0.8	0.62	0.54	0.73	0.64	0.64
6	0.77	0.55	0.5	0.72	0.62	0.61
7	0.74	0.49	0.46	0.71	0.59	0.59
8	0.71	0.43	0.43	0.7	0.57	0.57
9	0.68	0.38	0.4	0.69	0.55	0.55
10	0.64	0.34	0.37	0.68	0.53	0.52

Table 2. Exponential-based reliability TSE model and two existing models' validation by RMSE value [2,20]

Design life	Percentage Error (%)							
	Antenna		Transponder		Amplifier		Battery	
	Nadirah/ New	Shazana/ New	Nadirah/ New	Shazana / New	Nadirah / New	Shazana / New	Nadirah / New	Shazana / New
1	-22.6	-28.71	-17.2	-23.76	-19.4	-25.74	-19.4	-25.74
2	-25.6	-24.72	-15.6	-14.61	-20	-19.10	-20	-19.10
3	-28.7	-21.52	-13.8	-5.06	-20.7	-12.66	-20.7	-12.66
4	-31	-17.14	-11.9	5.71	-20.2	-4.29	-20.2	-4.29
5	-32.5	-12.90	-8.8	17.74	-20	3.23	-20	3.23
6	-35.1	-9.09	-6.5	30.91	-19.5	12.73	-20.8	10.91
7	-37.8	-6.12	-4.1	44.90	-20.3	20.41	-20.3	20.41
8	-39.4	0.00	-1.4	62.79	-19.7	32.56	-19.7	32.56
9	-41.2	5.26	1.5	81.58	-19.1	44.74	-19.1	44.74
10	-42.2	8.82	6.3	100.00	-17.2	55.88	-17.2	52.94
RMSE value	34.2	16.1	10.2	49.7	19.6	28.2	19.9	27.6

Table 2 shows the calculation of RMSE values to indicate the difference between the previous and the new TSE models for a ten-year design life. Each subsystem's errors are computed and benchmarked each year, and the last row gives the overall RMSE for every case in comparison. The individual RMSE value indicates that Nadirah's model has the least agreement with the new TSE model among all the subsystems, with RMSE values of 34.2 for the antenna, 49.7 for the transponder, 19.6 for the amplifier, and 19.9 for the battery. Meanwhile, Shazana's model seems to have a better alignment, mainly during the early years of design life, showing RMSE values of 16.1 for the antenna, 28.2 for the transponder, 28.2 for the amplifier, and 27.6 for the battery. Yet, as the design life increases, the error associated

https://doi.org/10.31436/iiumej.v26i3.3788

with Shazana's model for the transponder and amplifier subsystems also increases noticeably. The data demonstrate that the new TSE model is now more accurate in predicting changes in subsystem reliability, as supported by a low RMSE value across all components, with the same trend.

The results of the exponential-based reliability TSE model for a satellite system's antenna, transponder, amplifier, and battery components, shown as reliability prediction curves, are presented in Fig. 2 (a). The curves display the predicted reliability of every element over the ten-year design life period. This TSE model gradually shows that each component slowly experiences a decline in reliability as time goes on, which reflects the normal wear and tear from age and heavy use [4]. The antenna starts with moderate reliability, and its value decreases gradually, whereas the transponder keeps reliability at a higher level for a more extended period, suggesting better durability. The amplifier and the battery display similar behaviors, but in the early years, the battery stays slightly more reliable until it levels off with the amplifier around year 6. Contrarily, the previous Exponential model does not consider changes over time. The earlier-developed approach uses an independent value of multiple satellites' reliability to predict reliability, failing to demonstrate how it may change with time. However, the previous Exponential model approach is still functional when the satellites are arranged by the launch date [5]. By this approach, the Exponential model helps assess the predicted reliability of satellites by their age compared to others. It allows the Exponential model to compare the difference between satellites from the older and current versions [5]. However, the time-dependent approach in the TSE model provides a better view by illustrating how reliability keeps deteriorating continuously. It is beneficial for observing and analyzing how the satellite behaves and helping with planning during its entire operation.

Fig. 2(b) illustrates different trends in reliability predictions that emerge when the proposed exponential-based reliability TSE model is compared with the two models from Nadirah and Shazana [2,20]. A smooth and uninterrupted decrease is seen in reliability values on the new TSE model for each subsystem: the antenna, transponder, amplifier, and battery. This trend indicates that satellites lose performance over time, as is common, giving a reasonable picture of their reliability. In contrast, Nadirah's model maintains higher reliability rates throughout the design period, decreasing only slightly. This may indicate a promising trajectory that does not align with satellite aging trends. Meanwhile, Shazana's model predicts that reliability begins above average but soon drops after the midpoint of design life, with a noticeable change from year five onwards. Shazana's model appears to overstate the drop in reliability later. This new TSE model is superior to the previous ones because it accounts for the impact of failure rate and model error by incorporating design-dependent parameters. The new TSE model also obtains low RMSE values for all systems, demonstrating its accuracy in making predictions. The model's continuous decline ensures it is well-suited for anticipating long-term problems and supporting decisions in satellite upkeep and planning.

https://doi.org/10.31436/iiumej.v26i3.3788

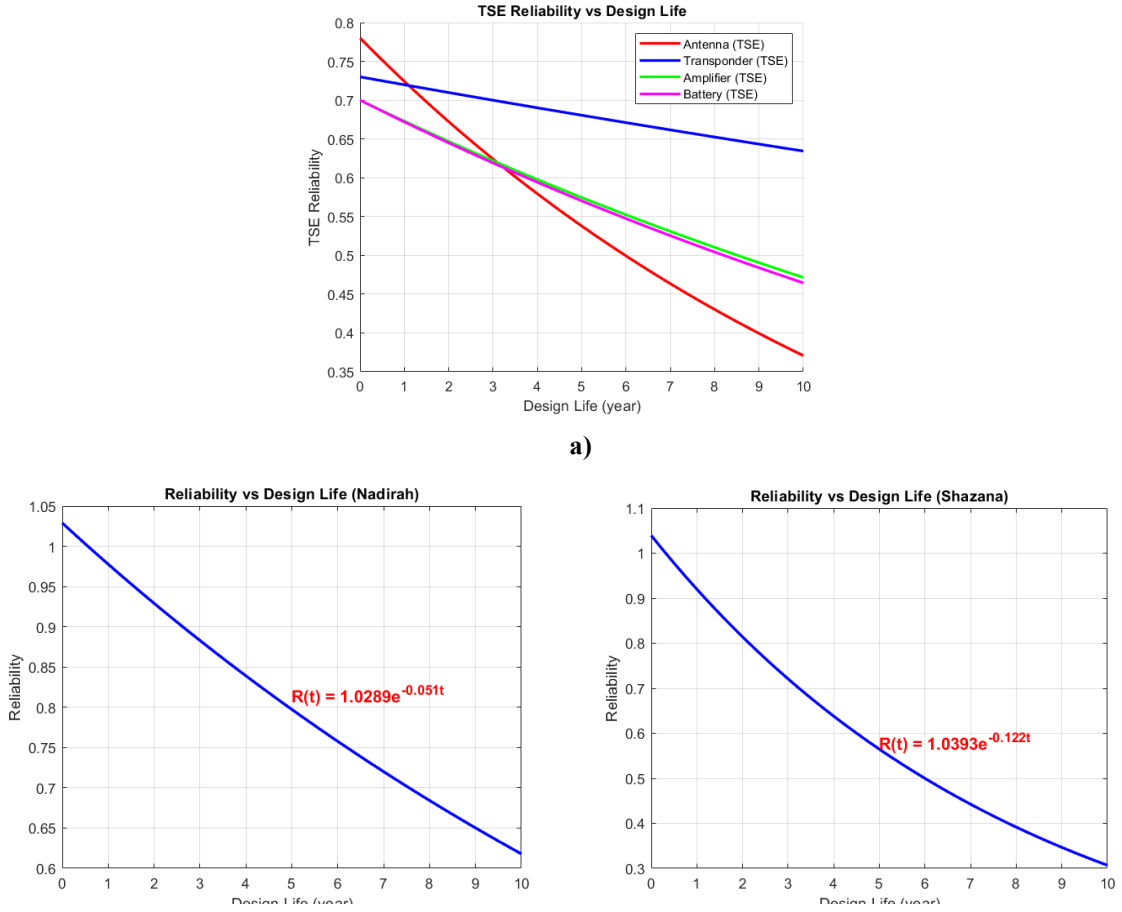


Figure 2. Reliability exponential-based TSE model prediction for each satellite anomaly against the two existing models (a) New Model, (b) Previous Models

5. CONCLUSION

The result of the new reliability TSE model developed by this research shows a promising outcome where RMSE values are 34.2, 10.2, 19.6, and 19.9 when compared with Nadirah's model and Shazana's model of 16.1, 49.7, 28.2, and 27.6 for antenna, transponder, amplifier, and battery, respectively. The model developed in this study provides a more accurate and realistic way to view the reliability of a satellite's subsystem than the previous models. By cooperating with failure rate, design life, and RMSE as parameters that vary with design, the TSE model accurately indicates a smooth and continuous reduction in reliability, which shows realistic subsystem performance. Before, most models remained reliable or decayed unrealistically. TSE shows better accuracy since the model consistently has low RMSE values in all subsystems. Because the model shows how performance or strength changes over time, it is helpful for preparations, design improvements, and assessing the product's lifetime performance. In the future, this model can be more accurate and valid by adding more subsystems to improve its predictive power. Moreover, considering environmental conditions, such as radiation and extreme temperature, the TSE reliability model can be more accurate in real-life situations. Therefore, the TSE approach can be used for maintenance scheduling, planning duplication, and assessing risk in past and future missions.

ACKNOWLEDGEMENT

This study was made possible with monetary assistance from the Asian Office of Aerospace Research and Development (AOARD) under grant scheme numbers FA2386-23-1-4073 and SPI23-179-0179. The authors would like to thank the IIUM CPS Khair Award for the assistance with tuition fees.

REFERENCES

[1] Reda, R., Ahmed, Y., Magdy, I., Nabil, H., Khamis, M., Refaey, A., Eldabaa, N., Elmagd, M. A., Lila, M. A., Ergawy, H., Elgarf, A., & Abed, G. (2023). Basic principles and mechanical considerations of satellites: A short review. Transactions on Aerospace Research, 2023(3), 40–54. https://doi.org/10.2478/tar-2023-0016.

[2] Rahman, N. S. A., & Rahim, N. A. (2023). Sustainable framework for a geostationary satellite control earth station system using parallel configuration. Indonesian Journal of Electrical Engineering and Computer Science, 30(3), 1498–1508.

[3] Diro, A., Kaisar, S., Vasilakos, A. V., Anwar, A., Nasirian, A., & Olani, G. (2024). Anomaly detection for space information networks: A survey of challenges, techniques, and future directions. Computers & Security, 139, 103705. https://doi.org/10.1016/j.cose.2024.103705

[4] Saleh, J. H., & Castet, J.-F. (2011). Spacecraft reliability and multi-state failures: A statistical approach. John Wiley & Sons. https://doi.org/10.1002/9781119994077

[5] Ahmad Mahmood, M. I. H., Rahim, N. A., & Mutholib, A. (2025, August 26–27). Satellite reliability analysis using mathematical prediction models: A case study of satellite anomalies. In International Conference on Computer and Communication Engineering (ICCCE), IIUM, Gombak, Kuala Lumpur.

[6] Rahim, N. A., Rahman, N. A., Abdoli, S., Rao, S., & Mokhtar, M. I. (2022). Analysis of failure frequency and failure rate of RF/antenna subsystems for an earth station system. NeuroQuantology, 20(10), 8654–8665.

[7] Rahman, N. S. A., & Rahim, N. A. (2022). Analysis of reliability prediction and maintainability activities of an earth station system using parallel configurations. In WCSE 2022 Spring Event: 2022 9th International Conference on Industrial Engineering and Applications (pp. 1647–1655).

[8] Aris, W. M. S. W., Rahim, N. A., Mansor, M. M., & Rahman, N. S. A. (2024). Operational cost models for an earth station system using 2-parallel and 4-parallel configurations. IIUM Engineering Journal, 25(2), 232–245.

[9] Muhammad, I., Wang, X., Li, C., Yan, M., & Chang, M. (2020). Estimation of the reliability of a stress–strength system from Poisson half logistic distribution. Entropy, 22(11), 1307. https://doi.org/10.3390/e22111307.

[10] Ezzati, G., & Rasouli, A. (2015). Evaluating system reliability using linear-exponential distribution function. International Journal of Advanced Statistics and Probability, 3(1), Article 3927. https://doi.org/10.14419/ijasp.v3i1.3927.

[11] Rahim, N. B. A., & Ferris, T. L. J. (2020). A method to establish a trade-space of system requirements and life cycle cost. IEEE Systems Journal, 14(1), 1257–1264.

[12] Almongy, H. M., Almetwally, E. M., Alharbi, R., Alnagar, D., Hafez, E. H., & Mohie El-Din, M. M. (2021). The Weibull generalized exponential distribution with censored sample: Estimation and application on real data. Complexity, 2021, Article 6653534. https://doi.org/10.1155/2021/6653534.

[13] Nagiya, K., & Ram, M. (2013). Reliability characteristics of a satellite communication system including earth station and terrestrial system. International Journal of Performability Engineering, 9, 667–676.

[14] Sedgwick, P. (2015). A comparison of parametric and non-parametric statistical tests. BMJ, 350, h2053. https://doi.org/10.1136/bmj.h2053.

[15] Wang, T., Chen, K., Zhang, Y., Zhang, K., Du, Z., & Xiao, F. (2021). Research on reliability evaluation method of protection system based on Kaplan-Meier and GO method. Journal of Physics: Conference Series, 1983, 012073. https://doi.org/10.1088/1742-6596/1983/1/012073

[16] Aris, W. S. M. W., & Rahim, N. A. (2023). Operational cost analysis of an earth station system using parallel configuration. In 2023 9th International Conference of Computer and Communication Engineering (ICCCE‘23) (pp. 388–393).

[17] Luengo, D., Martino, L., Bugallo, M., et al. (2020). A survey of Monte Carlo methods for parameter estimation. EURASIP Journal on Advances in Signal Processing, 2020(25). https://doi.org/10.1186/s13634-020-00675-6.

[18] Mutholib, A., Rahim, N. A., Gunawan, T. S., & Kartiwi, M. (2025). Trade-space exploration with data preprocessing and machine learning for satellite anomalies reliability classification. IEEE Access.

[19] Mutholib, A., Rahim, N. A., Gunawan, T. S., & Ahmarofi, A. A. (2024). Performance comparison of data preprocessing methods for trade-space exploration with AI model: Case study of satellite anomalies detection. In Proceedings of the IEEE 10th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA) (pp. 271–275). https://doi.org/10.1109/ICSIMA62563.2024.10675571.

[20] Rahim, N. A., & Nordin, N. (2020). Reliability model and proposed maintainability activities of earth station system. International Journal of Electrical, Electronics and Data Communication, 5, 20–24.

[21] Grile, T. M., & Bettinger, R. A. (2022). Statistical reliability estimation for satellites operating from 1991–2020 with payload reliability focus. In Proceedings of the 6th International Conference on System Reliability and Safety (ICSRS) (pp. 378–386).

https://doi.org/10.31436/iiumej.v26i3.3794

TRADE-SPACE EXPLORATION COMPARISON OF PARAMETRIC COST MODELS FOR SATELLITE ANOMALIES WITH RMSE AND RRMSE

MUHAMMAD ARYSS ZAFRAN SHAHRUL ARIFFIN[1], NADIRAH ABDUL RAHIM[1*], ABDUL MUTHOLIB[1,2]

[1] *Department of Electrical and Computer Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, 53100, Gombak, Kuala Lumpur, Malaysia*
[2] *Department of Information Systems, Faculty of Science and Technology, UIN Syarif Hidayatullah Jakarta, Banten 15412, Indonesia*

**Corresponding author: nadirahabdulrahim@iium.edu.my*

(Received: 15 June 2025; Accepted: 11 August 2025; Published online: 9 September 2025)

ABSTRACT: Satellites are essential for modern communication, navigation, and Earth observation, but their operation in the harsh space environment makes them vulnerable to anomalies and failures, resulting in significant financial losses. This research investigates the relationship between the design life of satellite subsystems and the cost of anomalies, using data from the Seradata database. Three statistical models, Exponential, Weibull, and Poisson distributions, were applied to cost data for four critical subsystems: antenna, payload, power system, and attitude control. Each model was fitted using one-variable (cost only) and two-variable approaches (design life and cost) to evaluate subsystems' behaviour independently and in combination. The performance of each model was assessed using the Root Mean Square Error (RMSE) and the Relative RMSE (RRMSE). Among all models, the Exponential distribution consistently produced the best results. For instance, the payload subsystem achieved the lowest RMSE and RRMSE of 115.73 and 41.66% in the two-variable case. Similarly, the attitude control subsystem showed strong performance with an RMSE of 110.57 and RRMSE of 40.59%. In contrast, the Poisson distribution yielded the highest errors across most subsystems, with the antenna subsystem reaching an RMSE of 489.87 and RRMSE of 102.65% in the two-variable fitting. The Weibull model demonstrated mixed performance, performing moderately for the payload subsystem (RMSE 186.21, RRMSE 67.03%), but poorly for the power system and antenna subsystems, particularly when two variables were used.

ABSTRAK: Satelit memainkan peranan penting dalam komunikasi moden, navigasi, dan pemerhatian bumi, namun operasinya dalam persekitaran angkasa lepas yang ekstrem menjadikannya terdedah kepada anomali dan kegagalan, sekali gus membawa kepada kerugian kewangan yang besar. Kajian ini mengkaji hubungan antara jangka hayat reka bentuk subsistem satelit dengan kos anomali menggunakan data daripada pangkalan data Seradata. Tiga model statistik iaitu taburan Eksponen, Weibull, dan Poisson telah digunakan pada data kos bagi empat subsistem kritikal: antena, muatan, sistem kuasa, dan kawalan sikap. Setiap model dipadankan menggunakan pendekatan satu pemboleh ubah (kos sahaja) dan dua pemboleh ubah (jangka hayat reka bentuk dan kos) bagi menilai tingkah laku subsistem secara individu dan gabungan. Prestasi setiap model dinilai menggunakan Ralat Punca Min Kuasa Dua (RMSE) dan Relatif RMSE (RRMSE). Antara semua model, taburan Eksponen secara konsisten memberikan dapatan terbaik. Sebagai contoh, dalam subsistem muatan, ia mencapai nilai RMSE dan RRMSE terendah iaitu 115.73 dan 41.66% masing-masing bagi kes dua pemboleh ubah. Begitu juga, subsistem kawalan sikap menunjukkan prestasi kukuh dengan RMSE 110.57 dan RRMSE 40.59%. Sebaliknya, taburan Poisson menghasilkan ralat tertinggi

https://doi.org/10.31436/iiumej.v26i3.3794

dalam kebanyakan subsistem, dengan subsistem antena mencatatkan RMSE 489.87 dan RRMSE 102.65% dalam padanan dua pemboleh ubah. Model Weibull pula menunjukkan prestasi bercampur, sederhana dalam subsistem muatan (RMSE 186.21, RRMSE 67.03%) tetapi lemah bagi subsistem sistem kuasa dan antena, khususnya apabila dua pemboleh ubah digunakan.

KEYWORDS: *Parametric cost mathematical models, satellite anomalies, trade-space exploration, Exponential, Weibull, Poisson, RMSE, RRMSE*

1. INTRODUCTION

Satellites are essential to modern communication, navigation, and Earth observation systems. However, operating in extreme space conditions makes them prone to anomalies and malfunctions, which can lead to substantial financial losses [1]. This research examines the relationship between failures in key satellite subsystems, including attitude control, power systems, payloads, and antennas, and their associated costs. By applying Trade-Space Exploration (TSE) in combination with statistical modelling, the study analyzes data sourced from the Seradata database [2]. Using MATLAB, cost models are developed using Exponential regression to capture compounding risk factors, Weibull distributions to represent failure rates over time, and Poisson distributions to model the frequency of discrete anomaly events. The central issue addressed is that many existing cost models overlook nonlinear cost escalation, the financial impact of insurance claims, and probabilistic predictions of anomalies [3]. Traditional models often oversimplify the complex trade-offs between design life, performance, and operational cost, making it difficult for satellite operators to make informed decisions [3]. Thus, this study addresses these shortcomings by integrating parametric cost modelling and evaluating model performance using Root Mean Square Error (RMSE) and Relative RMSE (RRMSE) to assess accuracy. Such an approach is crucial for enhancing space situational awareness and supporting efforts to minimize space debris from malfunctioning or failed satellites [4].

2. SATELITE SUBSYSTEMS FAILURES

Understanding subsystem vulnerabilities is key to analyzing satellite anomalies, as some components are more susceptible to failure than others, which often leads to major operational setbacks and financial implications. Data sourced from Seradata highlights four subsystems with the highest recorded anomaly rates: the power system, attitude control system, payload system, and antenna subsystem [2]. Each of these plays a critical role in the satellite's overall performance, and their tendency toward failure points to the need for targeted design improvements and robust monitoring methods.

The power system, which comprises solar panels, batteries, and distribution electronics, is crucial for sustaining satellite operations. Failures in this area can range from reduced solar panel efficiency to total battery malfunction [2, 5]. Solar panels often suffer damage from micrometeoroid impacts or gradually lose efficiency due to prolonged exposure to radiation. Batteries, meanwhile, are vulnerable to thermal cycling and overcharging, which can diminish their capacity or even cause catastrophic failure [6]. Notably, intense radiation events have been known to cause either temporary disruptions or irreversible harm to these power subsystems, significantly shortening mission duration [2, 5].

The attitude control system (ACS) maintains the satellite's proper orientation in space. It ensures that instruments and communication links are accurately directed and solar panels receive optimal sunlight. Failures within the ACS, especially in components like reaction

https://doi.org/10.31436/iiumej.v26i3.3794

wheels, can cause a loss of orientation, making it impossible for the satellite to perform essential operations [2, 6]. Reaction wheels are prone to mechanical wear and friction-based issues, particularly over extended periods, and failure in one or more of these components can compromise the mission. Because these problems are challenging to address once the satellite is in orbit, real-time telemetry and early-warning systems are critical for managing ACS health [7].

The payload system, which differs based on the satellite's function (such as imaging, communication, or scientific research), is essentially the heart of the mission. Payload failures may stem from various sources, including extreme temperature shifts, vibrations during launch, or radiation damage to sensitive instruments [5, 8]. In communication satellites, for example, amplifier degradation can weaken signal strength or disrupt data transmission altogether [2, 5].

Finally, the antenna subsystem sends and receives signals between the satellite and Earth. This system is vulnerable to mechanical issues like misalignment, wear and tear, and damage from space debris or radiation exposure [2, 9]. Misalignment can reduce communication efficiency, resulting in data loss or operational delays. Smaller satellites that rely on deployable antennas face additional risk; these antennas can malfunction during deployment, potentially crippling the satellite's communication capabilities [2, 5].

3. RESEARCH METHODOLOGY

This section presents the methodology adopted to explore the trade-space between satellite subsystem design life and anomaly-related costs using traditional mathematical modelling, to determine how varying design life values influence financial losses, and identify cost-effective configurations through a structured process encompassing problem identification, data gathering, visualization, model fitting, evaluation, and trade-space analysis.

3.1. Dataset

The dataset used in this study was obtained from the Seradata database, which documents satellite missions and their associated anomalies [2]. Only records involving satellites that experienced one or more subsystem anomalies were considered. Each data entry includes the design life and corresponding anomaly cost for four specific subsystems: antenna, payload, power system, and attitude control. About 87 satellite cost data points ranged from LEO, MEO, and GEO orbits. The cost values represent the total financial loss, including the satellite's manufacturing cost and any insurance claims resulting from the anomaly [2]. Table 1 tabulates the Seradata components. The values are concealed due to confidentiality reasons. Seradata advised the researchers not to reveal the values. The reason is that some satellite companies are not keen to reveal specific cost values.

Table 1. Seradata Components

Subsystems	Design Life (years)	Cost (US$m)
Antenna	15	XXX
	..	..
Payload	10	XXX
	..	..
Power System	5	XXX
	..	..
Attitude Control	13	XXX
	..	..

3.2. Parametric Cost Models

Based on the nature of the data and the goal of analyzing cost trends relative to reliability, three mathematical distributions commonly used in reliability engineering were selected: the Exponential, Weibull, and Poisson distributions. These models were chosen for their ability to capture the behaviour of failure rates and lifetime costs in complex systems [10, 11]. Each model was applied to the cost versus design life data of each subsystem separately to identify the distribution that best describes the cost pattern. The three models, namely the Exponential, Weibull, and Poisson, are defined in Eqs. (1), (2), and (3), respectively.

$$f(x) = ae^{-bx} \tag{1}$$

where x denotes the design life of the subsystem, a and b are the model parameters.

$$f(x) = abx^{(b-1)}e^{-(ax^b)} \tag{2}$$

where x denotes the design life of the subsystem, a is the cost scale, and b is the Weibull parameter.

$$f(x) = \frac{e^{-a}a^x}{x!} \tag{3}$$

where a denotes the rate parameter and x is the number of failures.

The modelling process was then done using MATLAB, which provided tools for curve fitting, parameter estimation, and visual validation. The design life and cost data for each subsystem align with the three selected models. The resulting model curves were plotted alongside the actual cost data to observe the level of agreement between predicted cost and observed cost values. MATLAB's built-in functions were employed to extract model parameters and evaluate the statistical quality of each fit.

3.3. Cost Models Performance

To assess the accuracy and effectiveness of each fitted model, two key error metrics were used: Root Mean Square Error (RMSE) and Relative Root Mean Square Error (RRMSE) [12]. RMSE measures the absolute difference between the predicted and actual cost values, indicating the model's prediction error [13]. RRMSE, on the other hand, expresses the error as a percentage relative to the mean of the observed values, offering a normalized measure that facilitates comparison across different subsystems or scales [14, 15]. The model with the lowest RMSE and RRMSE for a given subsystem was considered the best-fitting model. These metrics played a crucial role in identifying which distribution most accurately captured the cost-design life relationship [16, 17]. The RMSE and RRMSE equations can be defined in Eqs. (4) and (5).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Observed\ Cost - Predicted\ Cost)^2} \tag{4}$$

$$RRMSE = \frac{RMSE}{Mean\ of\ Actual\ Cost} \times 100 \tag{5}$$

3.4. Trade-Space Analysis

Following model fitting and evaluation, a trade-space analysis was conducted to assess the impact of variations in subsystem design life on the overall cost of anomalies [18, 19]. Both one-variable and two-variable fitting approaches were used during this process. The one-variable fitting focused on the relationship between design life and cost for individual subsystems, allowing for a clearer understanding of how each subsystem behaves

independently [17, 20]. In contrast, two-variable fitting considered the interaction between design life and cost across multiple subsystems simultaneously, offering a broader view of how combined subsystem behaviour influences overall cost trends. This approach was necessary to capture both isolated and interdependent effects on cost [21]. The analysis helped to identify critical points where extending the design life no longer leads to proportional cost savings, thus revealing optimal design zones [21, 22]. These insights contribute to making informed decisions on allocating design resources efficiently while balancing reliability and cost considerations [23, 24].

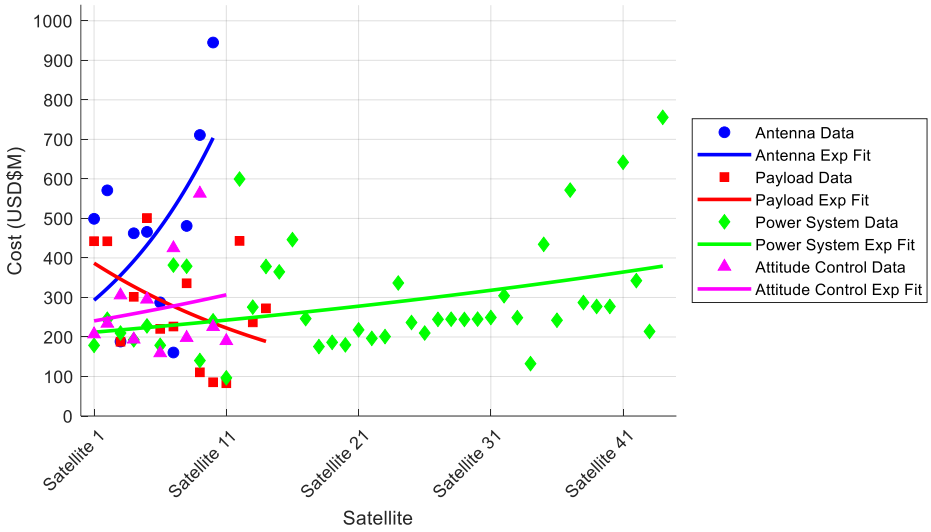
4. RESULTS

The graphical representations for Exponential, Weibull, and Poisson distributions, both one-variable and two-variable, can be found in Figures 1 and 2, respectively. From Figure 1, the cost data analyzed through Exponential curve fitting reflects the financial consequences of satellite subsystem anomalies. Among the subsystems, the antenna-related anomalies exhibited the steepest Exponential rise in cost, with values ranging from US$160.7 million to US$945 million, and an average of US$477.2 million. This sharp increase underscores the growing criticality and complexity of antenna systems, where even less frequent failures can result in extremely high financial losses. Such trends may be due to the critical role of antennas in communication and navigation, where in-orbit failures are difficult and costly to rectify [25]. In contrast, the payload subsystem showed a decreasing Exponential cost trend, with costs ranging from US$83 million to US$501 million, and an average of US$277.8 million. This decline may indicate technological maturity, standardization, and improved anomaly mitigation strategies in payload systems, resulting in a reduced financial impact over time, as designs have become more reliable, easier to test, and better equipped to handle minor issues before they escalate [25, 26].

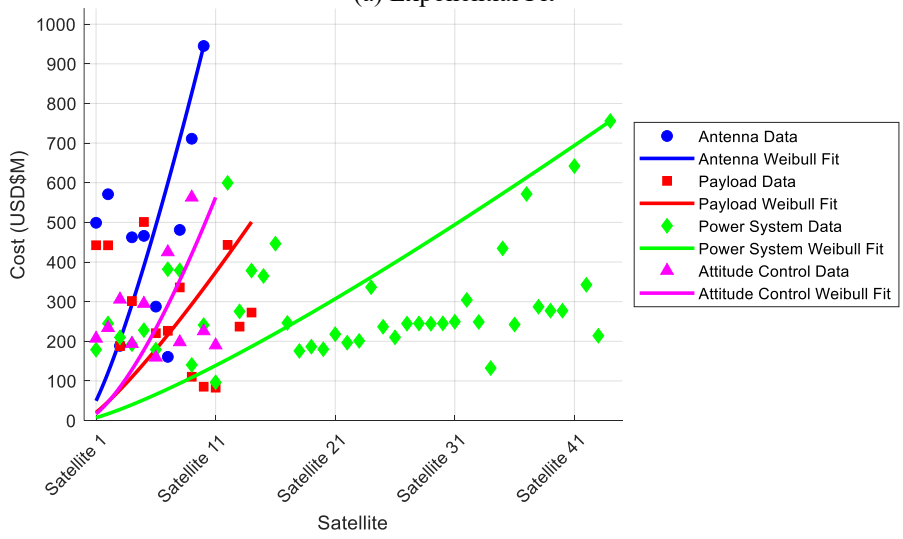
The power subsystem displayed a generally stable cost pattern, with values between US$96.6 million and US$755.75 million, and an average cost of US$288.2 million. Although most power-related anomalies result in moderate financial losses, a few significant spikes indicate that some failures, such as those involving solar arrays or batteries, can cause substantial economic damage. Finally, the attitude control subsystem reflected a moderate upward trend in cost, with anomaly-related expenses ranging from US$159.25 million to US$563 million, and an average of US$272.4 million. This suggests that while attitude control issues are generally manageable, the critical nature of this subsystem in maintaining satellite orientation and mission continuity makes its failures progressively more expensive to address.

For the Weibull fit, it can be deduced that the antenna subsystem shows a steep Weibull rise, reflecting the consequences of anomalies that are likely tied to the growing complexity and importance of antenna systems in satellite missions. The payload subsystem follows a less steep trajectory, with costs spread more evenly, pointing to improvements in payload resilience or mission-specific impact. Power systems exhibit a Weibull curve with a long tail, suggesting most anomalies incur moderate costs but outliers remain highly expensive, particularly in newer, power-intensive missions [27]. Attitude control systems demonstrate more stable cost behavior, yet still with a rising trend, highlighting the non-negligible effect of pointing failures on mission performance.

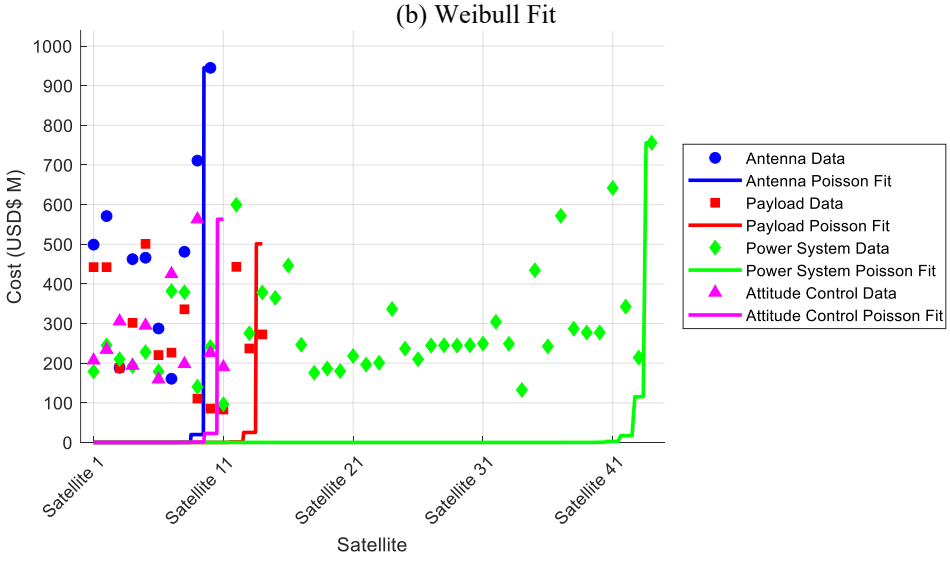
https://doi.org/10.31436/iiumej.v26i3.3794



(a) Exponential Fit



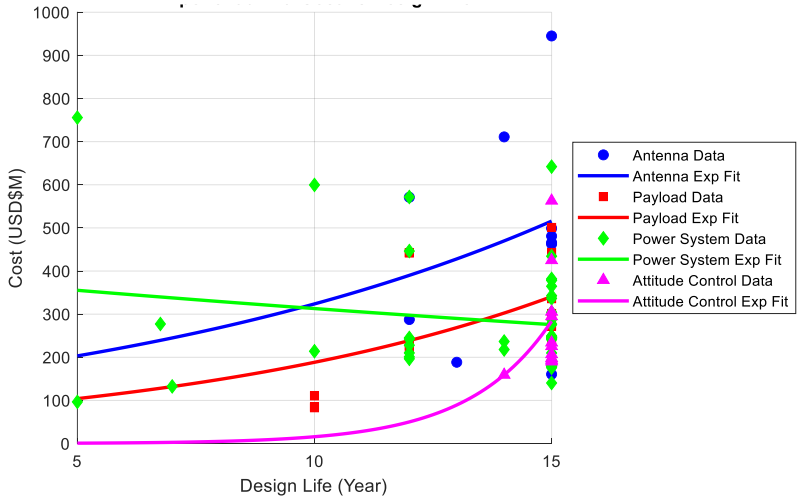
(b) Weibull Fit



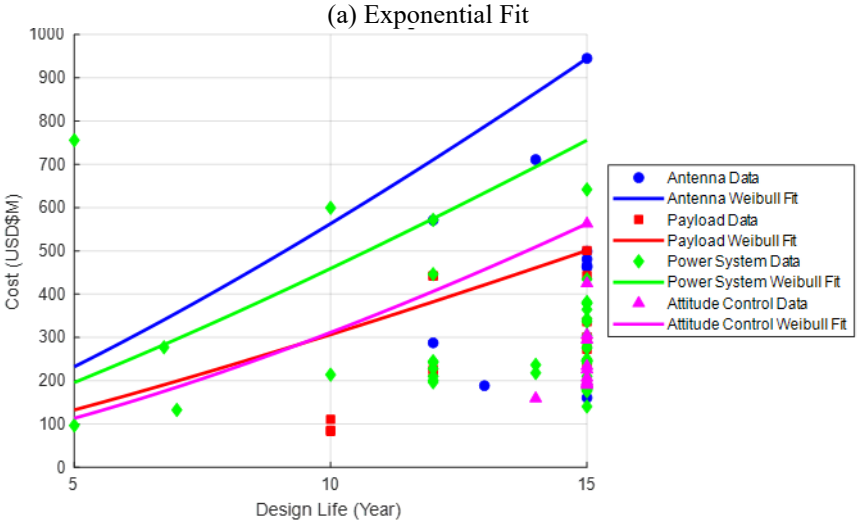
(c) Poisson Fit

Figure 1. Cost Data for Various Fits

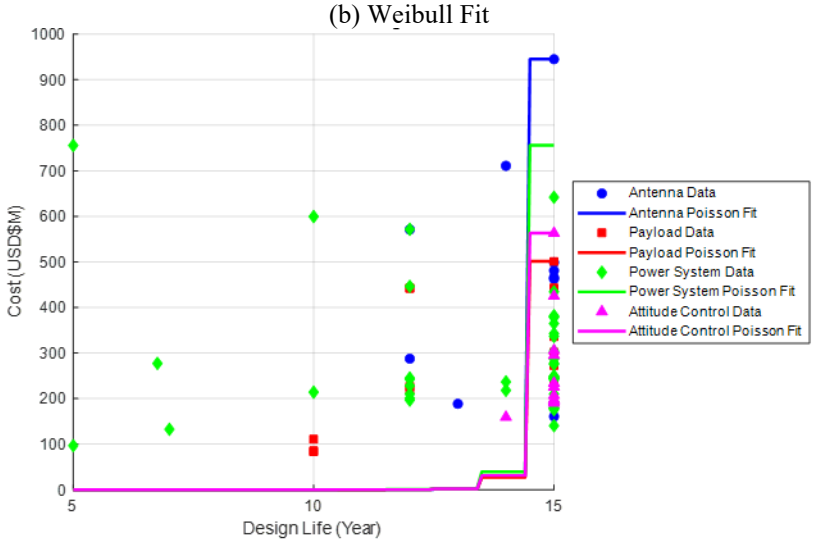
https://doi.org/10.31436/iiumej.v26i3.3794



(a) Exponential Fit



(b) Weibull Fit



(c) Poisson Fit

Figure 2. Cost vs Design Life for Various Fits

On the other hand, as shown in Figure 2, the Exponential model was applied to the cost versus design life data for each satellite subsystem. In Figure 2, 2 variables are used: design life and cost. The cost data is like a one-variable graph. For the antenna subsystem with a design life ranging from 12 to 15 years, the curve closely followed the upward trend in cost as design

https://doi.org/10.31436/iiumej.v26i3.3794

life increased, indicating a strong fit. The payload subsystem with values from 10 to 15 years also exhibited a reasonable Exponential increase, suggesting a consistent relationship between longevity and investment. In contrast, the power system's fitted curve unexpectedly decreased with design life, with values ranging between 5 and 15 years. This does not align with engineering expectations and suggests that this subsystem may not suit an Exponential model. The attitude control subsystem, which has values only in 14 and 15 years, displayed an exaggerated rise in cost, possibly due to limited or uneven data in the higher design life range. The Exponential model worked best for the antenna and payload subsystems but was less reliable for the power system and attitude control components.

The Weibull model was applied to the cost versus design life data to evaluate its flexibility in modeling various subsystem behaviors. For the antenna subsystem with an average design life of 13.5 years, the Weibull fit produced a smooth upward curve that effectively captured the increasing cost trend. The payload subsystem with values between 10 and 15 years also showed a strong alignment between the fitted curve and the actual data, suggesting a consistent cost increase as design life extended. The power system's compromise of design life between 5 and 15 years, which did not follow the exponential pattern well, was better represented by the Weibull fit, which provided a realistic, gradual rise in cost. Similarly, the attitude control subsystem with values of 14 and 15 years demonstrated improved curve fitting, with the Weibull model managing to capture the cost behavior more smoothly than the exponential approach. Overall, the Weibull model offered better adaptability across all subsystems and was particularly effective in modeling the power system and attitude control costs.

The Poisson model was tested to explore its applicability to the cost versus design life data; however, the results revealed significant limitations. For all subsystems, which are antenna, payload, power system, and attitude control, with average values of design life of 13.5 years, 12.5 years, 10 years, and 14.5 years, respectively, the fitted curves resembled discrete step functions rather than smooth trends. This characteristic of the Poisson model led to poor alignment with the actual data, especially for subsystems like power system and attitude control, where the cost varied more gradually. Since the Poisson distribution is designed to model count-based events rather than continuous variables, such as cost, its application was conceptually mismatched. As a result, the Poisson model failed to provide meaningful or realistic predictions and was the least suitable option for this analysis.

Table 2. RMSE and RRMSE results

Subsystems	**Model**	**RMSE (1 variable)**	**RRMSE (1 variable)**	**RMSE (2 variable)**	**RRMSE (2 variable)**
Antenna	Exponential	194.94	71.56%	218.16	45.72%
	Weibull	271.85	99.80%	455.24	95.40%
	Poisson	431.14	158.27%	489.87	102.65%
Payload	Exponential	121.84	44.73%	115.73	41.66%
	Weibull	239.07	87.76%	186.21	67.03%
	Poisson	304.71	111.86%	240.17	86.45%
Power System	Exponential	128.17	47.05%	134.70	46.73%
	Weibull	229.69	84.32%	434.17	150.64%
	Poisson	295.86	108.61%	442.79	153.63%
Attitude Control	Exponential	114.15	41.91%	110.57	40.59%
	Weibull	194.22	71.30%	307.02	112.71%
	Poisson	310.01	113.80%	290.91	106.79%

https://doi.org/10.31436/iiumej.v26i3.3794

After executing the models on the dataset, we evaluated their performance in predicting actual satellite subsystem costs. The Root Mean Square Error (RMSE) and Relative Root Mean Square Error (RRMSE) were used as the primary evaluation metric, as they quantify the average deviation between the predicted values and the observed data. A lower RRMSE indicates a better fit of the model to the actual cost values. Table 2 compares RMSE and RRMSE between all subsystems across the two types of graphs.

The Exponential distribution consistently provided the best fit among the models, especially when design life was used as a second variable [28]. This was particularly evident for the payload subsystem, where the Exponential model achieved the lowest RMSE and RRMSE values of 121.84 and 44.73% in the one-variable case, and even lower in the two-variable case at 115.73 and 41.66%, respectively. Similarly, for the attitude control subsystem, the Exponential model performed well, with an RMSE of 114.15 and an RRMSE of 41.91% in one-variable fitting, and showed slight improvement in the two-variable case, with an RMSE of 110.57 and an RRMSE of 40.59%.

On the other hand, the Poisson model consistently produced the highest errors across most subsystems. For instance, in the antenna subsystem, the Poisson model had an RMSE of 431.14 and RRMSE of 158.27% in the one-variable case, and even higher RMSE in the two-variable case at 489.87, though the RRMSE slightly decreased to 102.65%. This suggests that Poisson modelling does not effectively capture the relationship between the cost and design life of these components, likely due to its discrete nature and assumptions that don't align well with continuous cost behavior [29].

The Weibull model showed mixed performance. It performed moderately well for the payload subsystem, with RMSE and RRMSE of 239.07 and 87.76% (one-variable), and 186.21 and 67.03% (two-variable), indicating an improvement when a second variable is considered. However, for the power system, the Weibull model was less reliable in the two-variable case, with RMSE increasing from 229.69 to 434.17 and RRMSE jumping from 84.32% to 150.64%, showing it failed to capture the system's behaviour under more complex modeling. Overall, the Exponential model best represents the underlying relationship between subsystem design life and the associated costs, especially in the two-variable form. It offers the most consistent and lowest errors across multiple subsystems, making it a more reliable choice for cost modeling in satellite anomaly analysis.

5. CONCLUSION

This study focused on understanding how anomalies in critical satellite subsystems, including the antenna, payload, power system, and attitude control, relate to their design life and subsequent financial losses. Using data from the Seradata database, the research employed statistical modeling within a trade-space exploration framework to analyze cost behavior and failure patterns. The goal was to provide more accurate cost predictions and support better design decisions by considering both the technical and financial aspects of satellite performance. The exponential distribution consistently performed best in fitting the data among the models tested. It yielded the lowest RMSE and RRMSE values for most subsystems, particularly when design life and cost were combined in the two-variable fitting. This was particularly evident for the payload subsystem, where the exponential model achieved an RMSE of $115.73 million and an RRMSE of 41.66%, and for the attitude control subsystem, with an RMSE of $110.57 million and an RRMSE of 40.59%. These results indicate that the exponential model captures the trend of rising costs with subsystem aging more effectively than the other distributions. In comparison, the Poisson model performed poorly across all subsystems, with high RMSE and RRMSE values that indicated a weak fit. For instance, the

antenna subsystem reached an RMSE of $489.87 million and RRMSE of 102.65%, highlighting its limitations when applied to continuous cost data. The Weibull model yielded mixed results, performing reasonably well for the payload subsystem with an RMSE of $186.21 million and an RRMSE of 67.03%, but failed to provide reliable predictions for the power system and antenna subsystems, particularly in two-variable modeling, where errors increased. The findings highlight that choosing the right model is crucial when analyzing satellite reliability and anomaly-related costs. The exponential model is the most reliable and consistent, making it a practical choice for future studies and design planning. Ultimately, this study provides a valuable framework for optimizing subsystem design and managing risk by identifying where extending design life stops delivering proportional cost benefits.

ACKNOWLEDGEMENT

This research received monetary assistance from the Asian Office of Aerospace Research and Development (AOARD) under the grant scheme numbers FA2386-23-1-4073 and SPI23-179-0179. The authors appreciate the IIUM CPS Khair Award for the assistance with tuition fees.

REFERENCES

[1] Kodheli, O., et al. (2021). Satellite communications in the new space era: A survey and future challenges. IEEE Communications Surveys & Tutorials, 23(1), 70–109. https://doi.org/10.1109/comst.2020.3028247.

[2] Shahrul Ariffin, M. A. Z., Abdul Rahim, N., & Mutholib, A. (2025, August 26–27). Trade-space exploration analysis for a system incorporating parametric cost mathematical models: A case study of satellite anomalies. International Conference on Computer and Communication Engineering (ICCCE), Gombak, Kuala Lumpur.

[3] Gao, Y., Yue, L., & Penghui, S. (2021, August). Research status of typical satellite communication systems. 2022 20th International Conference on Optical Communications and Networks (ICOCN). https://doi.org/10.1109/icocn53177.2021.9563909.

[4] Hofner, K., Vahl, A., & Stoll, E. (2018, January). Cost-efficient satellite constellation design by network reliability analysis. 2022 Annual Reliability and Maintainability Symposium (RAMS), 1–6. https://doi.org/10.1109/ram.2018.8463029.

[5] Sakaguchi, K., & Nagatsuma, T. (2022). Impact of space environment on geostationary meteorological satellite data outage. Space Weather, 20(5). https://doi.org/10.1029/2021sw002965.

[6] Jin, W., Sun, B., Li, Z., Zhang, S., & Chen, Z. (2019). Detecting anomalies of satellite power subsystem via stage-training denoising autoencoders. Sensors, 19(14), 3216. https://doi.org/10.3390/s19143216.

[7] Riveiros, A. P., Xing, Y., Bastianello, N., & Johansson, K. H. (2024, June). Real-time anomaly detection and categorization for satellite reaction wheels. 2022 European Control Conference (ECC), 253–260. https://doi.org/10.23919/ecc64448.2024.10591184.

[8] Shen, R., et al. (2024, August). A dynamic channel emulator for CubeSat system test and integration. Proceedings, 2, 129–130. https://doi.org/10.1109/ccece59415.2024.10667238.

[9] Buzulukova, N., & Tsurutani, B. (2022). Space weather: From solar origins to risks and hazards evolving in time. Frontiers in Astronomy and Space Sciences, 9. https://doi.org/10.3389/fspas.2022.1017103.

[10] Poletto, J. (2022). An alternative to the exponential and Weibull reliability models. IEEE Access, 10, 118759–118778. https://doi.org/10.1109/ACCESS.2022.3219426.

[11] Mahmoudi, E., & Sepahdar, A. (2012). Exponentiated Weibull-Poisson distribution: Model, properties and applications. Mathematics and Computers in Simulation, 92, 76–97. https://doi.org/10.1016/j.matcom.2013.05.005.

[12] Nawaz, M., Haseeb, M., Qamar, S., Hassan, W., & Shahzad, A. (2024). Gene expression programming-based multivariate model for earth infrastructure: Predicting ultimate bearing capacity of rock socketed shafts in layered soil-rock strata. Modeling Earth Systems and Environment. https://doi.org/10.1007/s40808-024-02061-9.

[13] Mutholib, A., Abdul Rahim, N., Gunawan, T. S., & Ahmarofi, A. A. (2024, July). Performance comparison of data preprocessing methods for trade-space exploration with AI model: Case study of satellite anomalies detection. IEEE 10th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA), 271–275. https://doi.org/10.1109/ICSIMA62563.2024.10675571.

[14] Naimi, S., Ayoubi, S., Demattê, J., Zeraatpisheh, M., Amorim, M., & Mello, F. (2021). Spatial prediction of soil surface properties in an arid region using synthetic soil image and machine learning. Geocarto International, 37, 8230–8253. https://doi.org/10.1080/10106049.2021.1996639.

[15] Rahman, N. S. A., & Abdul Rahim, N. (2022). Analysis of reliability prediction and maintainability activities of an earth station system using parallel configurations. WCSE 2022 Spring Event: 2022 9th International Conference on Industrial Engineering and Applications, 1647–1655.

[16] Yang, Y., et al. (2022). Mapping crop leaf area index at the parcel level via inverting a radiative transfer model under spatiotemporal constraints: A case study on sugarcane. Computers and Electronics in Agriculture, 198, 107003. https://doi.org/10.1016/j.compag.2022.107003.

[17] Li, Y., et al. (2022). An enhanced spatiotemporal fusion method – Implications for DNN based time-series LAI estimation by using Sentinel-2 and MODIS. Field Crops Research. https://doi.org/10.1016/j.fcr.2022.108452.

[18] Muhammad W. A., & Abdul Rahim, N. (2023). Operational cost analysis of an earth station system using parallel configuration. 2023 9th International Conference on Computer and Communication Engineering (ICCCE), 388–393.

[19] Aris, W. M. S. W., Abdul Rahim, N., Mansor, M. M., & Rahman, N. S. A. (2024). Operational cost models for an earth station system using 2-parallel and 4-parallel configurations. IIUM Engineering Journal, 25(2), 232–245.

[20] Abdul Rahim, N., Rahman, N. A., Abdoli, S., Rao, S., & Mokhtar, M. I. (2022). Analysis of failure frequency and failure rate of RF/antenna subsystems for an earth station system. NeuroQuantology, 20(10), 8654–8665.

[21] Mutholib, A., Abdul Rahim, N., Gunawan, T. S., & Kartiwi, M. (2025). Trade-space exploration with data preprocessing and machine learning for satellite anomalies reliability classification. IEEE Access.

[22] Abdul Rahim, N. B. A., & Ferris, T. L. J. (2020). A method to establish a trade-space of system requirements and life cycle cost. IEEE Systems Journal, 14(1), 1257–1264.

[23] Rahman, N. S. A., & Abdul Rahim, N. (2023). Sustainable framework for a geostationary satellite control earth station system using parallel configuration. Indonesian Journal of Electrical Engineering and Computer Science, 30(3), 1498–1508.

[24] Abdul Rahim, N., & Nordin, N. (2020). Reliability model and proposed maintainability activities of earth station system. International Journal of Electrical, Electronics and Data Communication, 5, 20–24.

[25] Denig, W. F., Redmon, R. J., Rodriguez, J. V., & Allen, J. H. (2014). Book review: Satellite anomalies: Benefits of a centralized anomaly database and methods for securely sharing information among satellite operators. Space Weather, 12(8), 528–529. https://doi.org/10.1002/2014sw001100.

[26] Grile, T. M., & Bettinger, R. A. (2022, November). Statistical reliability estimation for satellites operating from 1991–2020 with payload reliability focus. 2022 6th International Conference on System Reliability and Safety (ICSRS), 378–386. https://doi.org/10.1109/icsrs56243.2022.10067366.

[27] Huguenin, M., et al. (2023, October). Statistical sizing of a satellite power subsystem. Proceedings, 1–6. https://doi.org/10.1109/espc59009.2023.10298153.

[28] Wang, B. (2006). Unbiased estimations for the exponential distribution based on step-stress accelerated life-testing data. Applied Mathematics and Computation, 173(2), 1227–1237. https://doi.org/10.1016/j.amc.2005.04.066.

[29] Bektashi, X., Rexhepi, S., & Limani–Bektashi, N. (2022). Dispersion of count data: A case study of Poisson distribution and its limitations. Asian Journal of Probability and Statistics, 18–28. https://doi.org/10.9734/ajpas/2022/v19i230464.

https://doi.org/10.31436/iiumej.v26i3.3800

OPTIMIZING GENERATION COST AND REDUCING GAS EMISSIONS IN POWER GENERATION USING THE ARTIFICIAL BEE RABBIT OPTIMIZATION ALGORITHM

WEI WEN LEE, MOHD RUZAINI BIN HASHIM*, CHIN KIM GAN

Faculty of Electrical Technology and Engineering, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia

**Corresponding author: ruzaini@utem.edu.my*

(Received: 20 June 2025; Accepted: 09 August 2025; Published online: 9 September 2025)

ABSTRACT: This study develops a hybrid metaheuristic optimization algorithm named Artificial Bee Rabbit Optimization (ABRO) to improve generation cost efficiency and reduce gas emissions in power generation systems. By integrating the strengths of the Artificial Bee Colony (ABC) and Artificial Rabbits Optimization (ARO) algorithms, ABRO aims to overcome issues such as premature convergence and slow convergence speed commonly observed in ABC and ARO. This paper evaluates and compares the ABRO algorithm against a collection of optimization algorithms, such as ABC, ARO, the Crow Search (CSA) algorithm, and the Artificial Jellyfish Search (JS) algorithm. The evaluation covers four benchmark functions and extends to engineering applications, specifically in solving the economic dispatch, emission dispatch, and an integrated objective that considers financial and emission dispatch aspects for the IEEE 26-bus system. The simulation results show that ABRO generally outperforms the competing algorithms tested in solving various benchmark functions. ABRO consistently achieved the lowest mean, standard deviation, and minimum values, demonstrating superior convergence speed, robustness, and accuracy. Furthermore, the ABRO algorithm effectively enhances optimization regarding generation cost, generation emission, and an integrated objective that considers both economic and emission dispatch aspects for the IEEE 26-bus system.

ABSTRAK: Kajian ini membangunkan satu algoritma pengoptimuman metaheuristik hibrid yang dinamakan Pengoptimuman Buatan Bee Rabbit (ABRO) bagi meningkatkan kecekapan kos penjanaan dan mengurangkan pelepasan gas dalam sistem penjanaan tenaga. Gabungan kekuatan antara Koloni Buatan Bee (ABC) dan Pengoptimuman Buatan Rabbit (ARO), menghasilkan ABRO yang bertujuan mengatasi isu penumpuan pramatang dan kelajuan penumpuan perlahan, ysng sering berlaku pada ABC dan ARO. Kajian ini menilai algoritma ABRO dan beberapa algoritma pengoptimuman lain seperti ABC, ARO, algoritma Pencarian Crow (CSA), dan pengoptimum Pencarian Buatan Jellyfish (JS). Penilaian meliputi empat fungsi penanda aras dan diperluas kepada aplikasi kejuruteraan, khususnya dalam penyelesaian masalah pengagihan ekonomi, pengagihan pelepasan, serta objektif bersepadu yang memgambil kira kedua-dua aspek ekonomi dan pelepasan bagi sistem IEEE 26-bas. Dapatan simulasi menunjukkan bahawa algoritma ABRO secara amnya mengatasi prestasi algoritma lain yang diuji dalam menyelesaikan pelbagai fungsi penanda aras. ABRO secara konsisten mencapai nilai min, sisihan piawai, dan nilai minimum terendah, sekali gus membuktikan kelajuan penumpuan, keteguhan, dan ketepatan terbaik. Tambahan pula, algoritma ABRO mampu meningkatkan pengoptimuman dari segi kos penjanaan, pelepasan penjanaan, serta objektif bersepadu yang mempertimbangkan kedua-dua aspek ekonomi dan pelepasan bagi sistem IEEE 26-bas.

https://doi.org/10.31436/iiumej.v26i3.3800

KEYWORDS: *Artificial Bee Colony, Optimization algorithm, Hybrid algorithm, Economic dispatch, Emission dispatch*

1. INTRODUCTION

Optimization is finding the optimal solution from a set of potential solutions. It either minimizes or maximizes the value of the objective function subject to some constraints [1]. Optimization algorithms are search methods that are designed to iteratively improve upon an initial solution until an optimal or near-optimal result is achieved [2],[3]. Artificial Bee Colony (ABC) [4], Artificial Jellyfish Search (JS) optimizer [5] and Artificial Rabbits Optimization (ARO) [6] are some examples of optimization algorithms that have been widely used in various fields for solving optimization problems.

According to the No-Free Lunch (NFL) theorem [7], there is no single optimization algorithm that can solve all problems well. ABC algorithm is an effective optimization approach for solving several practical issues, such as those in medical image processing [8], power dispatch [9] and wireless sensor networks [10]. However, the basic ABC algorithm undergoes a slow convergence rate and easily traps the local optima in specific scenarios [11]. Therefore, researchers have modified and adapted the ABC algorithm to suit specific problems and enhance its performance. Some of the variants of the ABC algorithm include hybrid ABC [12], adaptive ABC [13] and multiobjective ABC [14].

One of the recent optimization algorithms, called the Artificial Bee Rabbit Optimization (ABRO) algorithm [15], was first proposed in 2023 during a conference. The ABRO algorithm is a hybrid algorithm inspired by the ABC algorithm, which simulates the foraging behavior of honeybees and rabbits adapting to their natural environment. ABRO algorithm synthesizes these two models, aiming to achieve higher accuracy and convergence speed. The paper [15] was primarily focused on proposing the ABRO algorithm and comparing it against its predecessor algorithms, ABC and ARO, without employing statistical tools to evaluate its performance.

In this study, the ABRO algorithm is evaluated against other competitor algorithms, such as ABC, ARO, and the Crow Search Algorithm (CSA) [16], and Artificial Jellyfish Search (JS) optimizer [5]. The evaluation includes four benchmark functions with different landscapes and complexities. Moreover, these algorithms are used in the IEEE 26-bus system, and their performances for tackling the economic dispatch [17], emission dispatch [18] and combined economic with emission dispatch problems [19] are evaluated. Moreover, the ABRO algorithm has been validated statistically using the Wilcoxon Signed-Rank Test. This study aims to demonstrate that the ABRO algorithm, developed through hybridization, performs better in optimization, particularly in solving economic and emission dispatch issues in the IEEE 26-bus system.

The subsequent sections of the paper are structured as follows: Section 2 briefly explains the ABC and ARO algorithms. Section 3 highlights the inspiration, mathematical works, and flows of the ABRO algorithm. Section 4 presents ABRO's comparative results with other algorithms in optimizing four benchmark functions and the IEEE 26-bus system. Finally, Section 5 concludes the study's findings.

https://doi.org/10.31436/iiumej.v26i3.3800

2. ARTIFICIAL BEE COLONY AND ARTIFICIAL RABBITS OPTIMIZATION ALGORITHMS

The Artificial Bee Colony (ABC) algorithm is well-known for its simplicity and strong global search capabilities, and has been successfully applied to solve problems across various fields. However, several studies have highlighted that ABC is poor in exploitation and suffers from issues such as premature convergence, slow convergence, and getting trapped in local optima [11], [12]. On the other hand, the Artificial Rabbits Optimization (ARO) algorithm is a newer technique designed to improve optimization performance. Despite its innovative design, the ARO still exhibits limitations in its exploration capabilities, which may lead to suboptimal solutions and stagnation [20]. Therefore, the hybridization technique is introduced to combine the strengths of both algorithms and enhance performance.

Hybridization in optimization algorithms involves combining multiple techniques, such as combining different metaheuristic algorithms or integrating machine learning methods [21]. The goal of the hybridization technique is to develop a more powerful and efficient optimization approach. The Artificial Bee Rabbit Optimization (ABRO) algorithm was developed based on this hybridization concept. To gain insight into the development of the proposed ABRO algorithm, it is essential to study the prior algorithms, namely the ABC and ARO algorithms.

2.1. Artificial Bee Colony Algorithm

The Artificial Bee Colony Algorithm (ABC) algorithm [4] was developed by Dervis Karaboga in 2005, which draws inspiration from the activity of honey bees when searching for food. In ABC, food sources are representative of solutions to the addressed problem. Each food source produces nectar proportionally to fit the solution to the problem. The ABC algorithm comprises three kinds of searching agents, such as employed bees (EBs), onlooker bees (OBs), and scout bees (SBs). EBs are in charge of searching for food sources around their hives, OBs select food based on the information they share by EBs, and SBs will explore new food sources if the EBs' solution has not improved for a specific iteration or exceeds the predefined limit. The ABC algorithm terminates once it reaches the predefined number of iterations or discovers a satisfied solution [22]. The ABC algorithm has demonstrated its applicability in various engineering applications, such as path planning of mobile robots, designing of optimized PID controllers, and MPPT power extraction [11]. The ABC algorithm is characterized by its simplicity, with few control parameters; however, it has limitations, including slow convergence and susceptibility to premature convergence. Hence, many researchers modify the ABC algorithm to enhance its performance, overcome its limitations, and adapt it to specific problems [12],[23].

Recent studies have shown that hybridizing the ABC algorithm with other techniques enhances its performance. For example, Yan et al. (2020) proposed a hybrid ABC with Tabu-Neighborhood search to improve route optimization in logistics [24], while Mala et al. (2023) introduced ABC variants (NMABC and LHABC) enhanced with neural network strategies and stochastic gradient descent for improved classification accuracy [25]. Moreover, Kumar et al. (2022) integrated ABC with Differential Evolution and Whale Optimization Algorithm (ABC-DE-WOA) to accelerate convergence [26]. Ramachandran et al. (2022) developed FRCSA-ABC by merging ABC with a fuzzy logic-enhanced Crow Search Algorithm for economic dispatch problems [27]; and Ustun et al. (2022) incorporated DE's mutation and crossover into a modified ABC for higher precision [28]. These works illustrate the growing interest in hybridizing ABC to expand its applicability and effectiveness.

https://doi.org/10.31436/iiumej.v26i3.3800

2.2. Artificial Rabbits Optimization

The Artificial Rabbits Optimization (ARO) algorithm [6] is inspired by the survival tactics employed by rabbits. The survival strategies of rabbits include energy shrink strategies, detour foraging techniques, and random hiding techniques. The detour foraging acts as the exploration phase, where the rabbits seek food near other rabbits' nests, which can prevent predators from discovering their nests. In contrast, random hiding acts as the exploitation phase, where the rabbits will randomly pick one of their burrows to hide in to avoid being captured by predators. The energy shrink approach relies on the energy factor, which leads to a transition from the detour foraging approach to the random hiding approach. The ARO algorithm continues to iterate until the stopping criterion is met, such as maximum iterations or the desired fitness value is achieved. The ARO algorithm has been extensively applied in various areas, such as PV allocation [29], stock price prediction [30] and water productivity prediction [31].

The ARO algorithm is a modern optimization approach that is easy to implement and requires minimal control parameters [6]. However, on the contrary, the ARO suffers from several shortcomings, including premature convergence, a dearth of population diversity, and an imbalance between exploitation and exploration [32]. For instance, to improve the efficiency of the ARO algorithm, researchers [32] have combined the ARO algorithm with the dimension learning-based hunting technique and have employed it to investigate optimal sizing of a stand-alone hybrid system. Recently, Li et al. (2024) proposed a hybrid algorithm combining ARO with the Bottlenose Dolphin Optimizer (BDO) to enhance feature selection in network intrusion detection tasks. Their modified algorithm, LBARO, incorporates multiple improvements such as an adaptive switching mechanism, Levy flight strategy, and dynamic lens imaging to overcome ARO's stagnation and improve exploration and exploitation. [33]. The success of LBARO across benchmark functions and classification datasets demonstrates the growing interest and effectiveness of hybrid ARO strategies in complex optimization problems.

3. METHODOLOGY

3.1. Fundamentals of Artificial Bee Rabbit Optimization

The Artificial Bee Rabbit Optimization (ABRO) algorithm [15] is an optimization algorithm that combines the techniques from two other algorithms, namely the ABC algorithm [4] and the ARO algorithm [6]. The ABRO algorithm is a metaheuristic approach that iteratively generates and evaluates candidate solutions to guide the search toward better solutions for various optimization problems. The ABRO algorithm was developed based on the observations made on two existing algorithms: ABC and ARO. ABC shows strong exploration and requires few control parameters, but it struggles with exploitation and converges slowly [34]–[36]. In contrast, ARO demonstrates rapid convergence and strong exploitation capabilities [6], but it is a relatively new algorithm proposed in 2022 and might have limitations in terms of performance, efficiency, and robustness.

Drawing inspiration from the strengths and limitations of both ABC and ARO, the ABRO algorithm has been developed. The aim is to achieve a balanced compromise between exploration and exploitation, using the strength of both algorithms and aiming to achieve better optimal outcomes. The ABRO algorithm combines the two key approaches from its predecessor algorithm. It involves the employed bee phase of the ABC and the random hiding strategy from the ARO, as shown in Figure 1.

https://doi.org/10.31436/iiumej.v26i3.3800

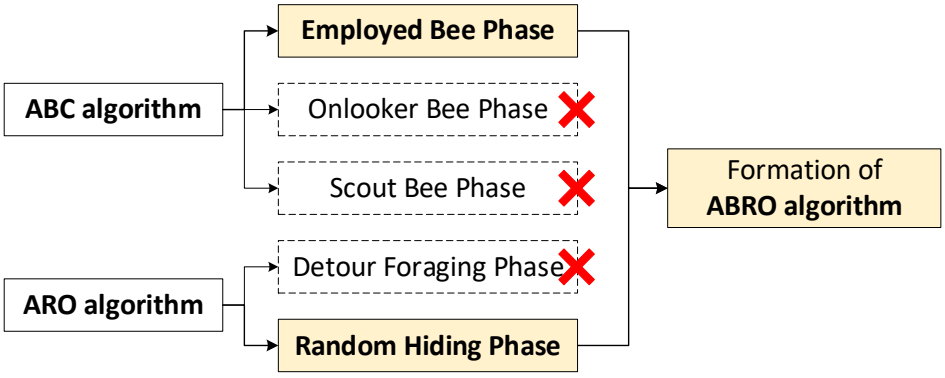


Figure 1. The formation of the ABRO algorithm

3.2. Mathematical Models of Artificial Bee Rabbit Optimization

The overview and the mathematical models of the ABRO are outlined. The ABRO algorithm consists of the employed bee stage and the random hiding stage. The total population is denoted NP refers to the combined count of search agents that are actively engaged in both stages. The term "food sources" (FS) relates to half of the population and refers to the search agents involved in either the employed bee phase or the random hiding phase. Each search agent is restricted to operating with a single food source or solution at a time. D represents the dimension or the number of input variables the optimization task takes.

The term Sol is an abbreviation used to denote a candidate solution or a feasible solution. During the initialization phase, these solutions are distributed at random throughout D-dimensional search space. Equation (1) is used to execute the initialization process.

$$Sol_{i,j} = rand(0,1) \times \left(ub_{i,j} - lb_{i,j}\right) + lb_{i,j} \tag{1}$$

$Sol_{i,j}$ represents the solution randomly found within the population, where i refers to the i^{th} food source with $i = 1,2,3,\ldots,FS$ and j refers to the j^{th} dimensional vector with $j = 1,2,\ldots,D$. The term $rand(0,1)$ means a pseudorandom number is generated in the interval of 0 to 1. The "ub" indicates the upper limit, where the function's largest possible value can attain. Conversely, the "lb" referred to as the lower bound, represents the minimum possible value that the function can take.

After the initialization process, the main loop of the ABRO algorithm starts. The method initiates with the employed bee stage. At this stage, each search agent visits its solution $Sol_{i,j}$ and evaluates its objective function value, which is denoted as $f(Sol_{i,j})$. During the employed bee phase, the search agent generates new solutions, $NSol_{i,j}$ based on its current candidate solution using equation (2) and evaluates its fitness using equation (3).

$$NSol_{i,j} = Sol_{i,j} + \left(Sol_{i,j} - Sol_{k,j}\right) \times \emptyset \tag{2}$$

where the k and j are random selection index, $k = 1,2,\ldots,FS$ but the k must be different from i whereas $j = 1,2,\ldots,D$. $\emptyset$ refers to the random number that is generated between -1 to 1. The fitness equation described in (3) is used to direct the search process. Xs appears as the variable or placeholder, which is used generically to represent either $Sol_{i,j}$ or $NSol_{i}$ can be inserted into the equation (3) depending on the context.

$$Fit(Xs_i) = \begin{cases} \frac{1}{1+f(Xs_i)} & ; \quad f(Xs_i) \geq 0 \\ 1 + |f(Xs_i)| & ; \quad f(Xs_i) \leq 0 \end{cases} \tag{3}$$

https://doi.org/10.31436/iiumej.v26i3.3800

During the search process, it is essential to ensure that the solutions generated by the algorithm remain within the defined search space. Hence, the new solution is checked against the defined ub and lb by using the equation in (4). If the generated solution's parameter exceeds these boundaries, it will be updated to fit within the limits.

$$NSol_{i,j} = \begin{cases} ub_{i,j} & ; \quad NSol_{i,j} > ub_{i,j} \\ lb_{i,j} & ; \quad NSol_{i,j} < lb_{i,j} \\ NSol_{i,j} & ; \quad lb_{i,j} < NSol_{i,j} < ub_{i,j} \end{cases} \tag{4}$$

Then, the search agents compare the fitness of the solutions and perform greedy selection using the equation as shown in (5). If the newly generated solution, $NSol_i$ exhibits greater fitness (Fit) than the existing candidate solution ($Sol_{i,j}$), the new solution, $NSol_i$ will be substituted and memorized for the next stage. If not, the existing candidate solution Sol_i remains.

$$Sol_i = \begin{cases} NSol_i & ; \quad Fit(NSol_i) > Fit(Sol_i) \\ Sol_i & ; \quad Fit(NSol_i) \leq Fit(Sol_i) \end{cases} \tag{5}$$

The subsequent phase is the random hiding phase. The variable "motion length" (L) represents the step length and can be calculated using equation (6). L is varied and will be generated from the higher to the lower steps along with the iterations. Higher steps enable the search agent to search in further locations, while the lower step is vice versa. T refers to the maximum number of cycles and $iter$ represents the iteration at a particular time.

$$L = \left(e - e^{\left(\frac{iter-1}{T}\right)^2} \right) \times sin(2\pi \times rand(0,1)) \tag{6}$$

The vector g as shown in equation (7) illustrates the random ordering of rd integers selected randomly from 1 to D without repeating. The term "$ceil$" in equation (8) represents the ceiling function that rounds the specific element to the nearest integer larger than or equal to that element. The g creates a random permutation of rd unique integers selected randomly from 1 to D.

$$g = randperm(D, rd) \tag{7}$$

$$rd = ceil(rand(0,1) \times D) \tag{8}$$

The vector c acts as the mapping vector and is used to randomly select the number of search agents' variables for mutation. The equation in (9) can be used to determine the vector c. The v represents the selected dimension vectors with possibly $v = 1, \ldots, [random(0,1) \times D]$ whereas p represents the p^{th} element of the search agents with $p = 1, \ldots, D$. R is the running operator that controls the running characteristic of search agents, and it can be calculated using equation (10).

$$c(p) = \begin{cases} 1 & if\ p == g(v) \\ 0 & else \end{cases} \tag{9}$$

$$R = L \times c \tag{10}$$

The random hiding stage is inspired by the survival technique of rabbits, digging different burrows and randomly choosing one of them as their nest for hiding, with the hope of confusing the predator and reducing the chances of being caught. Equation (11) selects which burrow/dimension vector to mutate randomly. Each search agent has several dimensional vectors, also known as variables. By adopting the hiding technique of rabbits, one of the dimensional vectors is selected for mutation.

https://doi.org/10.31436/iiumej.v26i3.3800

$$gr(q) = \begin{cases} 1 & if\ q == ceil(rand(0,1) \times D) \\ 0 & else \end{cases} \quad (11)$$

where q represents the only chosen dimension vector with possibly $q = 1, \dots, D$. gr represents the search agent's dimension vectors (burrows), if $gr = 1$ represents the selected burrow/ dimension vector to mutate and if $gr = 0$ represents the selected dimension vector is not affected.

H as shown in equation (12) represents the hiding parameter, and it will reduce linearly from 1 to $1 / T$ iteratively. The search agent starts searching in broader neighbourhoods and narrows down as iterations increase. b as shown in equation (13) mutate the randomly selected dimension vector/burrow.

$$H = (T - iter + 1)/T \quad (12)$$

$$b_i = Sol_i + H \times gr \times Sol_i \quad (13)$$

Subsequently, by using equation (14), the newly generated solution $NSol_i$ is generated. The i^{th} search agent intends to vary its current solution toward the randomly selected dimension vector (burrow) where i= $1, \dots, FS$. The boundaries of $NSol_i$ are checked to make sure they are within the search space.

$$NSol_i = Sol_i + R \times (rand(0,1) \times b_i - Sol_i) \quad (14)$$

Finally, the solution update for the ith search agent is determined using equation (15). The selection is based on the objective function value $f()$. In the minimization problem, if the newly generated solution $NSol_i$ has a lesser objective function value than the existing solution Sol_i, the solution will be updated to the newly generated solution $NSol_i$.

$$Sol_i = \begin{cases} NSol_i & ; \quad f(Sol_i) > f(NSol_i) \\ Sol_i & ; \quad f(Sol_i) \leq f(NSol_i) \end{cases} \quad (15)$$

Table 1. Pseudocode of ABRO algorithm

	ABRO Algorithm
1:	Set the parameters. NP, FS, D, T, ub, lb
2:	Initialize the population of candidate solutions of FS using (1)
3:	$iter$=1
4:	**While** $iter \leq T$ **do**
5:	**Employed Bee Stage:**
6:	***for all*** *i = 1: FS* ***do***
7:	Utilize (2) to create new solutions
8:	Use (3) to perform a fitness evaluation for the solutions
9:	Analyse and update the solutions using (5)
10:	***end for***
11:	**Random Hiding Stage:**
12:	***for all*** *i = 1: FS* **do**
13:	Determine running operator R using (6 – 10)
14:	Calculate b using (11 – 13)
15:	Generate a new solution using (14)
16:	Evaluate the objective function and do selection using (15)
17:	Update the search agent's solution
18:	***end for***
19:	Memorize the best solution and the best function value
20:	Iteration = Iteration+1
21:	**Until** ($iter = T$)

https://doi.org/10.31436/iiumej.v26i3.3800

The ABRO algorithm repeats the process by iteratively refining its solution in the search space until it meets the optimal solution or reaches the predefined maximum iterations. Table 1 shows the pseudocode of the ABRO algorithm. The pseudocode includes key components such as the initialization, iteration steps, employed bee stage, random hiding stage, and termination conditions.

4. RESULTS AND DISCUSSION

Based on previous studies, the ABC algorithm has limitations, including slow convergence and a tendency to get stuck at local optima. To address these shortcomings, the ABRO algorithm has been developed to improve ABC's convergence speed and accuracy. To evaluate the algorithm's effectiveness, the ABRO algorithm's results are compared with those of four existing algorithms: ABC, ARO, JS, and CSA, across four benchmark functions and in solving economic and emission dispatch problems for the IEEE 26-bus system. The ABRO algorithm demonstrates faster convergence and superior performance compared to the others. Further details are provided in this section.

4.1. Benchmark Functions

The ABRO algorithm has been evaluated with four benchmark functions, including the Sum Squares function, Rastrigin function, Ackley function, and Zakharov function that are taken from the literature [5]. Benchmark functions are sets of functions suitable for evaluating the performance of any optimization tasks. Several existing optimizers, such as ABC, ARO, CSA, and JS, are used to compare with ABRO. Each optimizer is run for 30 independent runs. To maintain equitable conditions for each algorithm, the number of search agents (NP) is adjusted to 50, whereas the maximum iteration (T) is selected to 100. Besides NP and T, ABC and CSA algorithms require additional parameter settings. The limit of ABC is set to $(NP/2) \times D$, while CSA requires setting the awareness probability to 0.1 and its flight length to 2. The optimal solution ($Best$), mean ($Mean$) and standard deviation (SD) of all algorithms are recorded in Table 2. Based on Table 2, the ABRO algorithm outperforms the other algorithms as it can achieve the best minimum value, the lowest SD and lowest $Mean$ in every selected benchmark function.

Wilcoxon Signed-Rank Test (WSRT) [37] has been utilized to validate the proposed ABRO's performance compared to competitor algorithms such as ABC, ARO, CSA, and JS across four benchmark functions. This statistical test is well-suited for assessing the significance of differences between paired data sets. The null hypothesis is that the median of differences between competitor algorithms and the proposed ABRO equals 0. The p-value reflects the probability of the null hypothesis being true. A significant level of 0.05 has been chosen. When the p-value is less than 0.05, it indicates statistically significant differences, leading to the rejection of the null hypothesis. "+" indicates significant improvement, "=" means no difference, and "−" refers to significant worsening in the ABRO algorithm versus competitors. Referring to Table 3, ABRO significantly improves over every benchmark function compared to other studied algorithms.

Figure 2 shows the convergence plots. ABRO converges faster than other algorithms in every benchmark function, as ABRO requires fewer iterations to find the optimal solution. Faster convergence is often desirable in optimization algorithms, especially when time and resources are valuable considerations. This efficiency can significantly reduce computational time and resources.

https://doi.org/10.31436/iiumej.v26i3.3800

Table 2. Optimization of benchmark mathematical functions with different algorithms

FUNCTION	INDEX	**ABRO**	ABC	ARO	JS	CSA
$f_1(x)$	*Mean*	**5.73E-17**	1.41E+02	2.47E-08	4.14E-04	7.75E+01
	SD	**1.72E-16**	1.51E+02	1.28E-07	2.11E-04	1.56E+01
	Best	**6.65E-30**	5.46E+00	2.60E-13	1.04E-04	4.74E+01
$f_2(x)$	*Mean*	**1.78E-15**	9.18E+01	1.11E-08	1.55E+02	1.16E+02
	SD	**4.83E-15**	1.56E+01	2.55E-08	1.81E+01	2.18E+01
	Best	**0.00E+00**	4.93E+01	7.46E-14	1.04E+02	7.71E+01
$f_3(x)$	*Mean*	**3.94E-09**	1.36E+01	1.41E-05	1.44E-02	6.38E+00
	SD	**9.28E-09**	1.16E+00	2.88E-05	2.51E-03	5.48E-01
	Best	**4.00E-15**	1.10E+01	3.73E-07	9.45E-03	5.33E+00
$f_4(x)$	*Mean*	**4.08E-15**	3.30E+01	1.23E-08	7.80E-02	5.42E+00
	SD	**1.54E-14**	1.50E+01	3.36E-08	6.13E-02	4.88E+00
	Best	**3.95E-22**	5.47E+00	4.80E-13	1.01E-02	7.22E-01

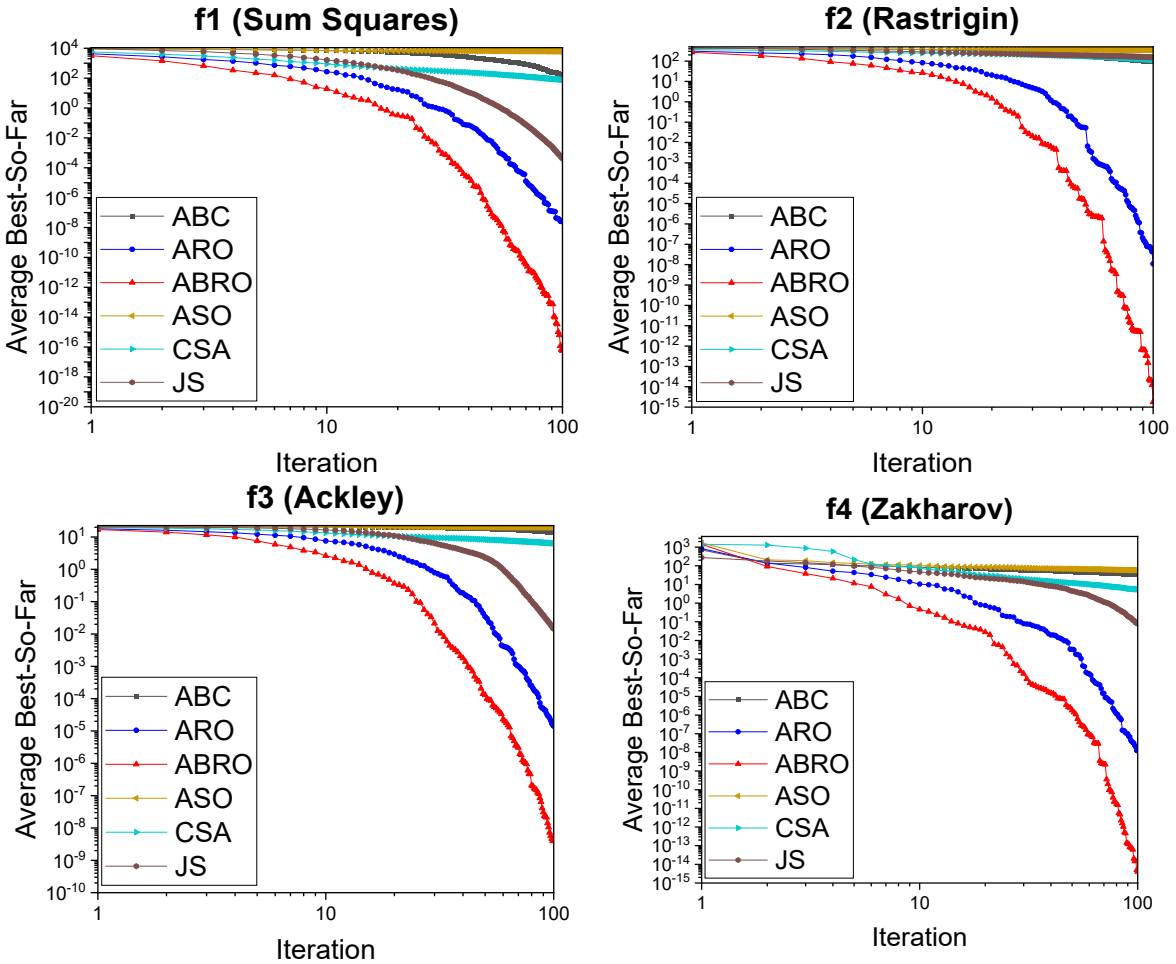


Figure 2. Comparisons of convergence curves

https://doi.org/10.31436/iiumej.v26i3.3800

Table 3. Significant comparison of WSRT for ABRO versus other competitors

Function	ABC versus ABRO			
	p-value	Median		Improve
		ABC	ABRO	
$f_1(x)$	2.00E-6	1.02E02	2.89E-19	+
$f_2(x)$	2.00E-6	9.19E+01	0.00E+00	+
$f_3(x)$	2.00E-6	1.39E+01	4.68E-10	+
$f_4(x)$	2.00E-6	3.35E+01	8.83E-17	+

	ARO versus ABRO			
	p-value	Median		Improve
		ARO	ABRO	
$f_1(x)$	2.00E-06	2.18E-10	2.89E-19	+
$f_2(x)$	2.00E-06	1.21E-10	0.00E+00	+
$f_3(x)$	2.00E-06	4.10E-06	4.68E-10	+
$f_4(x)$	2.00E-06	4.68E-10	8.83E-17	+

Function	JS versus ABRO			
	p-value	Median		Improve
		JS	ABRO	
$f_1(x)$	2.00E-06	3.64E-04	2.89E-19	+
$f_2(x)$	2.00E-06	1.61E+02	0.00E+00	+
$f_3(x)$	2.00E-06	1.37E-02	4.68E-10	+
$f_4(x)$	2.00E-06	5.94E-02	8.83E-17	+

	CSA versus ABRO			
	p-value	Median		Improve
		CSA	ABRO	
$f_1(x)$	2.00E-06	7.88E+01	2.89E-19	+
$f_2(x)$	2.00E-06	1.16E+02	0.00E+00	+
$f_3(x)$	2.00E-06	6.37E+00	4.68E-10	+
$f_4(x)$	2.00E-06	3.83E+00	8.83E-17	+

4.2. EEE 26-Bus System

Next, the ABRO algorithm and four other algorithms are implemented on the IEEE 26 bus system [38] to evaluate their applicability in optimizing the economic dispatch, emission dispatch, and combined economic and emission dispatch problems. The IEEE 26-bus system is a widely used test system in power systems research and optimization. The IEEE 26-bus system is a small-scale system that comprises five generators and one slack bus [38]. The generators are located respectively at buses 2, 3, 4, 5, and 26, with different power generation ranges [39] as described in Table 4. The system's power demand is 1,263MW. Each generator in the IEEE 26-bus system has its unique cost and emission coefficients, as listed in Table 4 [40]. Optimization's aim in this context is typically to determine the optimal allocation of power generation among accessible generating units in a power system to meet power demand, considering various constraints and objectives.

This study will primarily concentrate on three cases: economic dispatch, emission dispatch, and the combined optimization of financial and emission dispatch. Each algorithm has been run 30 times for every case to evaluate the consistency. Also, each algorithm has been configured with an NP value of 10 and a T value of 100. The recorded metrics include the mean, standard deviation, minimum, and maximum value.

Table 4. Generator limits, cost coefficients, and emission coefficients for the IEEE 26-bus system

Generator P_{Gi}	Power limits megawatt (MW)	Cost Coefficient			Emission Coefficient				
		a_i	b_i	c_i	α_i	β_i	γ_i	ε_i	λ_i
P_{G1} (Slack)	$100 \le P_{G1} \le 500$	0.0070	7.0	240	4.091	-5.543	6.490	2.0e-4	2.857
P_{G2}	$50 \le P_{G2} \le 200$	0.0095	10.0	200	2.543	-6.047	5.638	5.0e-4	3.333
P_{G3}	$80 \le P_{G3} \le 300$	0.0090	8.5	220	4.258	-5.094	4.586	1.0e-6	8.000
P_{G4}	$50 \le P_{G4} \le 150$	0.0090	11.0	200	5.326	-3.550	3.380	2.0e-3	2.000
P_{G5}	$50 \le P_{G5} \le 200$	0.0080	10.5	220	4.258	-5.094	4.586	1.0e-6	8.000
P_{G26}	$50 \le P_{G26} \le 120$	0.0075	12.0	190	6.131	-5.555	5.151	1.0e-5	6.667

https://doi.org/10.31436/iiumej.v26i3.3800

4.2.1. Case 1: Economic Dispatch

Economic dispatch aims to minimize the overall cost of generation while meeting the load demand by identifying the optimal schedule of all generating units. Economic dispatch is essential because it can help reduce the operating costs of power systems, leading to lower electricity prices for consumers. Its mathematical model is formulated in equation (16). In a power system with N generators are actively engaged in operation, the total generation cost can be expressed as FC, where FC is the sum of the individual generation costs of each generator. P_{Gi} represents the output power from the generator, while a_i, b_i, and c_i are the cost coefficients of the i^{th} generation unit in the IEEE 26-bus system.

$$FC(\vec{P}_G) = \sum_{i=1}^{N} a_i P_{Gi}^2 + b_i P_{Gi} + c_i \qquad dollar/hr \tag{16}$$

Table 5 lists the algorithms' performances in reducing the generation cost for the IEEE 26-bus system. ABRO consistently exhibits the lowest mean and the smallest standard deviation and attains the lowest minimum values compared to the other algorithms. ABRO can achieve mean generation costs of $ 15446.7425 per hour. Also, the convergence curve of all the algorithms for minimizing the generation costs is shown in Figure 3.

Table 5. Comparisons of Generation costs for the IEEE 26-bus (case 1)

Algorithm	Cost (dollars/hr)			
	Mean	SD	Min (Best)	Max (Worst)
ABRO	**15446.7425**	**0.0008**	**15446.7407**	**15446.7437**
ABC	15447.0838	0.4662	15446.7894	15448.6256
ARO	15447.1746	0.1897	15446.9767	15447.8997
JS	15446.8690	0.1229	15446.7697	15447.2949
CSA	15449.9793	1.8234	15448.5454	15455.4806

Furthermore, the WSRT has been run, and the output indicates that the ABRO algorithm significantly outperforms the other competitors for economic dispatch in the IEEE 26-bus system. To illustrate this point, the median ABRO, $Mdn = 15446.74$ is statistically significantly lower than the median ABC, $Mdn = 15446.84$, $z = -4.78$, $p = 2.00E-06$. Besides, the median ABRO, $Mdn = 15446.74$ is statistically significantly lower than the median ARO, $Mdn = 15447.14$, $z = -4.78$, $p = 2.00E-06$. Also, the median ABRO, $Mdn = 15446.74$ is statistically significantly lower than the median JS, $Mdn = 15446.82$, $z = -4.78$, $p = 2.00E-06$. Lastly, the median ABRO, $Mdn = 15446.74$ is statistically significantly lower than the median CSA, $Mdn = 15449.30$, $z = -4.78$, $p = 2.00E-06$.

4.2.2. Case 2: Emission Dispatch

The second case comprises emission dispatch. Emission dispatch focuses on environmental factors and aims to minimize the emission of pollutants, such as nitrogen oxides, sulphur dioxide, and carbon dioxide. By scheduling the generating units effectively, emission dispatch can help reduce emissions ton/hr while maintaining the power supply. Its mathematical model is represented in equation (17). γ_i, β_i, α_i, ε_i and λ_i are the emission coefficients of atmospheric pollutants.

$$EM(\vec{P}_G) = \sum_{i=1}^{N} 10^{-2} (\gamma_i P_{Gi}^2 + \beta_i P_{Gi} + \alpha_i) + \varepsilon_i e^{\lambda_i P_{Gi}} \qquad ton/\text{hr} \tag{17}$$

Table 6 lists the algorithms' performance in reducing the generation emission for the IEEE 26-bus system. In Table 6, ABRO consistently demonstrates the lowest mean and the smallest standard deviation, achieving the best minimum value when compared to the other algorithms. The ABRO algorithm can achieve an average generation emission of 15869.7847 tons/hr. In

https://doi.org/10.31436/iiumej.v26i3.3800

addition, the convergence curve of all the algorithms for generation emission in the IEEE 26 bus system is illustrated in Figure 3.

Table 6. Comparisons of generation emission for the IEEE 26-bus (case 2)

Algorithm	Emissions (ton/hr)			
	Mean	SD	Min (Best)	Max (Worst)
ABRO	**15869.7847**	**0.0001**	**15869.7846**	**15869.7852**
ABC	15870.0631	1.5252	15869.7846	15878.1385
ARO	15977.2962	68.2082	15896.8664	16142.6559
JS	15881.3237	3.9165	15871.8956	15887.0627
CSA	16704.0726	297.5323	15982.4369	17159.0742

Moreover, the WSRT has been run, and the output indicates that the ABRO algorithm significantly outperforms the other competitors except the ABC algorithm for emission dispatch in the IEEE 26-bus system. To illustrate this point, the median ABRO, Mdn=15869.78 is statistically significantly lower than the median ARO, $Mdn = 15955.43$, $z = -4.78$, $p = 2.00E-06$. Also, the median ABRO, $Mdn = 15869.78$ is statistically significantly lower than the median JS, $Mdn = 15881.50$, $z = -4.78$, $p = 2.00E-06$. In addition, the median ABRO, $Mdn = 15869.78$ is statistically significantly lower than the median CSA, $Mdn = 16754.61$, $z = -4.78$, $p = 2.00E-06$. Lastly, although ABRO's emission results are better on average than ABC's, the WSRT reveals no significant difference (p-value = 0.6). This suggests that while ABRO generally performs better, the difference is not statistically significant at the 0.05 level.

4.2.3. Case 3: Economic with Emission Dispatch

The third case is about solving the objective of encompassing the economic and emission dispatch problem. The combined optimization of economic and emission dispatch is a power system optimization problem that determines the optimal generation schedule for generators in the system, minimizing both the total generation cost and total emissions of pollutants. To address the dual objectives of cost minimization and emission reduction into a single objective, the weighted sum approach is used [41]. The weighted sum method can be represented as in (18). F refers to the objective function, p refers to the number of objectives ($p = 1,2, \dots, k$), w_p refer as the real-value weighting with $0 \leq w_p \leq 1$ and $\sum w_p = 1$.

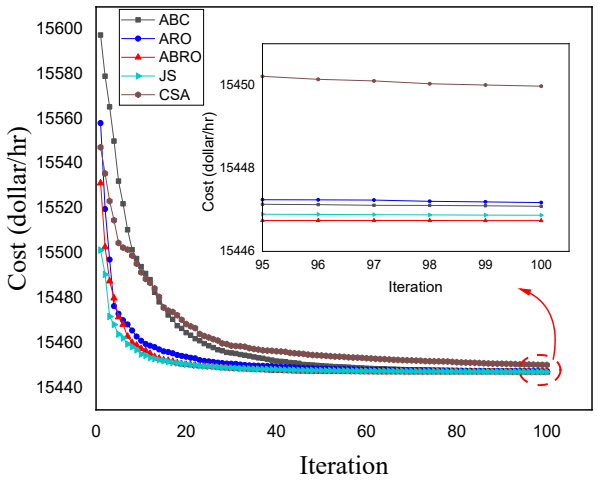
$$F = \sum_{p=1}^{k} w_p F_p \tag{18}$$

The weighted sum-based method converts the multi-objective optimization problem into a single-objective optimization problem. In this method, weights are assigned to each objective, and the optimal solution is obtained by linearly combining the respective objective function values. In the third case, two primary goals are considered: reducing generation costs and reducing generation emissions. Both objectives are considered equally important; hence, each is allocated a weight of 0.5, essentially treating them equally.

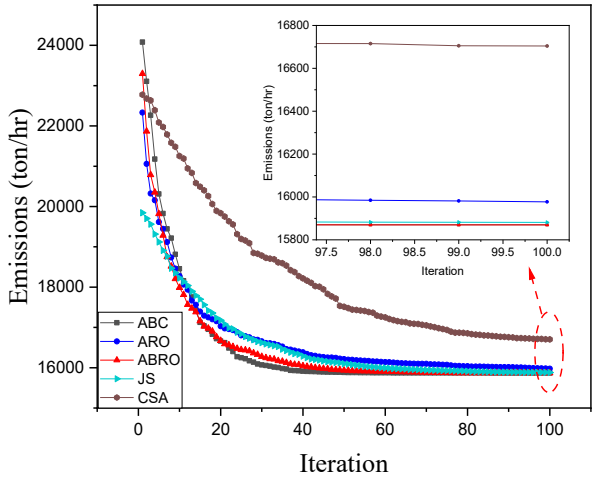
Table 7. Comparisons of generation costs with emissions for the IEEE 26-bus (case 3)

Algorithm	Costs with Emissions			
	Mean	SD	Min (Best)	Max (Worst)
ABRO	**15749.0905**	**9.25E-12**	**15749.0905**	**15749.0905**
ABC	15749.4444	1.85E+00	15749.0905	15759.2313
ARO	15797.1685	2.88E+01	15762.7517	15876.3179
JS	15761.5552	1.02E+01	15751.6845	15801.6833
CSA	16128.5023	1.16E+02	15898.6975	16318.4441

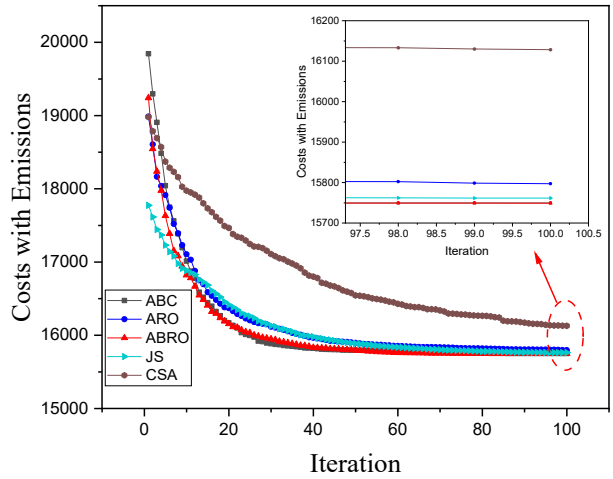
https://doi.org/10.31436/iiumej.v26i3.3800



(a) Case 1



(b) Case 2



(c) Case 3

Figure 3. Convergence curves for the IEEE 26-bus system

The outcomes of addressing economic and emission considerations using a weighted sum approach are presented in Table 7. In Table 7, ABRO consistently demonstrates the lowest

mean (15749.0905) and the smallest SD, and successfully finds the best Min compared to the other algorithms in optimizing the generation costs with emission. The WSRT is conducted, revealing that the ABRO algorithm significantly outperforms its competitors for weighted sum economic and emission dispatch on the IEEE 26-bus system. Specifically, the median ABRO, $Mdn = 15749.09$ is statistically significantly lower than the median ARO, $Mdn = 15793.84$, $z = -4.78$, $p = 2.00E - 06$. Besides, the median ABRO, $Mdn = 15749.09$ is statistically significantly lower than the median JS, $Mdn = 15759.08, z = -4.78, p = 2.00E - 06$. Also, the median ABRO, $Mdn = 15749.09$ is statistically significantly lower than the median CSA, $Mdn = 16156.96$, $z = -4.78$, $p = 2.00E - 06$. Lastly, the ABRO is statistically significantly lower than CSA with $z = -4.78$ and $p = 2.00E - 06$. Figure 3 illustrates the convergence curves for three cases: generation costs, generation emissions, and combined generation costs and emissions. As shown in Figure 3, ABRO demonstrates its superior convergence performance and achieves the lowest mean value at iteration 100 for these three cases.

5. CONCLUSION

This study introduces ABRO, a hybrid algorithm that combines the employed bee and random hiding phases from ABC and ARO. Hybridization significantly improved ABRO's performance, where ABRO demonstrated improved accuracy and convergence speed in finding solutions. This research evaluated ABRO's performance against ABC, ARO, CSA, and JS across four benchmark functions and the IEEE 26-bus system. ABRO consistently outperformed the others, achieving the lowest mean, standard deviation, and minimum values, demonstrating its robustness and efficacy. Applied to the IEEE 26-bus system, ABRO reduced generation costs, emissions, and their combined impact, surpassing other algorithms. Using the ABRO algorithm, this study achieved an average rate of 15,446.7425 dollars/hr for economic dispatch, 15,869.7847 tons/hr for emission dispatch, and 15,749.0905 for combined economic and emission dispatch. WSRT confirmed ABRO's superiority except in emission dispatch, where it did not significantly outperform ABC. However, this research was limited to four benchmark functions and the IEEE 26-bus system. Future studies could explore additional benchmark functions and applications. Moreover, integrating Pareto optimization techniques into ABRO could enhance its ability to solve multi-objective problems, resulting in a set of non-dominated solutions representing the trade-offs between conflicting objectives.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Centre for Research and Innovation Management (CRIM), UTeM, and Universiti Teknikal Malaysia Melaka (UTeM) for funding this research paper.

REFERENCES

[1] Mahmoud M. El-Halwagi, Ed., “Overview of optimization,” in process system engineering, vol. 7, Academic Press, 2006, pp. 285–314. doi: 10.1016/S1874-5970(06)80012-3.

[2] C. Peel and T. K. Moon, “Algorithms for Optimization [Bookshelf],” IEEE Control Syst., vol. 40, no. 2, pp. 92–94, Apr. 2020, doi: 10.1109/MCS.2019.2961589.

[3] A. P. Engelbrecht, Computational Intelligence, Second. Wiley, 2007. doi: 10.1002/9780470512517.

[4] D. Karaboga, “An idea based on honey bee swarm for numerical optimization., Technical Report - TR06” in Erciyes University, 2005, pp. 1–10.

[5] J.-S. Chou and D.-N. Truong, "A novel metaheuristic optimizer inspired by behavior of jellyfish in ocean," Appl. Math. Comput., vol. 389, p. 125535, Jan. 2021, doi: 10.1016/j.amc.2020.125535.

[6] L. Wang, Q. Cao, Z. Zhang, S. Mirjalili, and W. Zhao, "Artificial rabbits optimization: A new bio-inspired meta-heuristic algorithm for solving engineering optimization problems," Eng. Appl. Artif. Intell., vol. 114, p. 105082, Sep. 2022, doi: 10.1016/j.engappai.2022.105082.

[7] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," IEEE Trans. Evol. Comput., vol. 1, no. 1, pp. 67–82, Apr. 1997, doi: 10.1109/4235.585893.

[8] Ş. Öztürk, R. Ahmad, and N. Akhtar, "Variants of Artificial Bee Colony algorithm and its applications in medical image processing," Appl. Soft Comput., vol. 97, p. 106799, Dec. 2020, doi: 10.1016/j.asoc.2020.106799.

[9] M. Ettappan, V. Vimala, S. Ramesh, and V. T. Kesavan, "Optimal reactive power dispatch for real power loss minimization and voltage stability enhancement using Artificial Bee Colony Algorithm," Microprocess. Microsyst., vol. 76, p. 103085, Jul. 2020, doi:

[10] Z. Wang, H. Ding, B. Li, L. Bao, and Z. Yang, "An Energy Efficient Routing Protocol Based on Improved Artificial Bee Colony Algorithm for Wireless Sensor Networks," IEEE Access, vol. 8, pp. 133577–133596, 2020, doi: 10.1109/ACCESS.2020.3010313.

[11] A. Sharma, A. Sharma, S. Choudhary, R. K. Pachauri, A. Shrivastava, and D. Kumar, "A review on artificial bee colony and its engineering applications," J. Crit. Rev., vol. 7, no. 11, pp. 4097–4107, 2020, doi: 10.31838/jcr.07.11.558.

[12] M. Zhang, Y. Pan, J. Zhu, and G. Chen, "ABC-TLBO: A Hybrid Algorithm Based on Artificial Bee Colony and Teaching-Learning-Based Optimization," in 2018 37th Chinese Control Conference (CCC), IEEE, Jul. 2018, pp. 2410–2417. doi: 10.23919/ChiCC.2018.8483829.

[13] A. Yurtkuran and E. Emel, "An adaptive artificial bee colony algorithm for global optimization," Appl. Math. Comput., vol. 271, pp. 1004–1023, Nov. 2015, doi: 10.1016/j.amc.2015.09.064.

[14] J. Luo, Q. Liu, Y. Yang, X. Li, M. Chen, and W. Cao, "An artificial bee colony algorithm for multi-objective optimisation," Appl. Soft Comput., vol. 50, pp. 235–251, Jan. 2017, doi: 10.1016/j.asoc.2016.11.014.

[15] W. W. Lee and M. R. Bin Hashim, "A Hybrid Algorithm Based on Artificial Bee Colony and Artificial Rabbits Optimization for Solving Economic Dispatch Problem," in 2023 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), IEEE, Jun. 2023, pp. 298–303. doi: 10.1109/I2CACIS57635.2023.10193351.

[16] A. Askarzadeh, "A novel metaheuristic method for solving constrained engineering optimization problems: Crow search algorithm," Comput. Struct., vol. 169, pp. 1–12, Jun. 2016, doi: 10.1016/j.compstruc.2016.03.001.

[17] M. A. Bagherian and K. Mehranzamir, "A comprehensive review on renewable energy integration for combined heat and power production," Energy Convers. Manag., vol. 224, p. 113454, Nov. 2020, doi: 10.1016/j.enconman.2020.113454.

[18] M. Gent and J. Lamont, "Minimum-Emission Dispatch," IEEE Trans. Power Appar. Syst., vol. PAS-90, no. 6, pp. 2650–2660, Nov. 1971, doi: 10.1109/TPAS.1971.292918.

[19] B. Dey, S. K. Roy, and B. Bhattacharyya, "Solving multi-objective economic emission dispatch of a renewable integrated microgrid using latest bio-inspired algorithms," Eng. Sci. Technol. an Int. J., vol. 22, no. 1, pp. 55–66, Feb. 2019, doi: 10.1016/j.jestch.2018.10.001.

[20] H. Jia et al., "Improved artificial rabbits algorithm for global optimization and multi-level thresholding color image segmentation," Artif. Intell. Rev., vol. 58, no. 2, 2025, doi: 10.1007/s10462-024-11035-3.

[21] M. Abdel-Basset, L. Abdel-Fatah, and A. K. Sangaiah, "Metaheuristic Algorithms: A Comprehensive Review," in Computational Intelligence for Multimedia Big Data on the Cloud

with Engineering Applications, Elsevier, 2018, pp. 185–231. doi: 10.1016/B978-0-12-813314-9.00010-4.

[22] J. Yao Lee, M. Ruzaini Hashim, and M. O. Tokhi, "Artificial Bee Colony Optimization Algorithm with Flexible Manipulator System," vol. 3, no. 2, pp. 75–82, 2020.

[23] N. B and J. J, "Hybrid ABC-BAT Optimization Algorithm for Localization in HWSN," Microprocess. Microsyst., p. 104024, Jan. 2021, doi: 10.1016/j.micpro.2021.104024.

[24] Y. Yan, J. Chen, H. Huang, and S. Yang, "A hybrid algorithm for three-dimensional loading capacitated vehicle routing problems with time windows," in Proceedings - 2020 7th International Conference on Information Science and Control Engineering, ICISCE 2020, L. S., D. Y., M. J., and C. Y., Eds., Xiamen University, School of Informatics, Xiamen City, China: Institute of Electrical and Electronics Engineers Inc., 2020, pp. 1534–1540. doi: 10.1109/ICISCE50968.2020.00304.

[25] C. Mala, V. Deepak, S. Prakash, and S. Lashmi Srinivasan, "A Hybrid Artificial Bee Colony Algorithmic Approach for Classification Using Neural Networks," in EAI/Springer Innovations in Communication and Computing, H. A., R. A., M. S., and C. M.-Y., Eds., National Institute of Technology, Tiruchirappalli, Tamil Nadu, India: Springer Science and Business Media Deutschland GmbH, 2021, pp. 339–359. doi: 10.1007/978-3-030-47560-4_28.

[26] P. Kumar, B. J. Sowmya, A. Kanavalli, and A. Vincent, "ABC-DE-WOA: A New Hybrid Algorithm for Optimization Problems," in 2022 4th International Conference on Circuits, Control, Communication and Computing (I4C), 2022, pp. 501–506.

[27] M. Ramachandran, S. Mirjalili, M. Malli Ramalingam, C. A. R. Charles Gnanakkan, D. S. Parvathysankar, and A. Sundaram, "A ranking-based fuzzy adaptive hybrid crow search algorithm for combined heat and power economic dispatch," Expert Syst. Appl., vol. 197, 2022, doi: 10.1016/j.eswa.2022.116625.

[28] D. Ustun, A. Toktas, U. Erkan, and A. Akdagli, "Modified artificial bee colony algorithm with differential evolution to enhance precision and convergence performance," Expert Syst. Appl., vol. 198, 2022, doi: 10.1016/j.eswa.2022.116930.

[29] M. Elshahed, M. A. Tolba, A. M. El-Rifaie, A. Ginidi, A. Shaheen, and S. A. Mohamed, "An Artificial Rabbits' Optimization to Allocate PVSTATCOM for Ancillary Service Provision in Distribution Systems," Mathematics, vol. 11, no. 2, p. 339, Jan. 2023, doi: 10.3390/math11020339.

[30] B. Gülmez, "Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm," Expert Syst. Appl., vol. 227, p. 120346, Oct. 2023, doi: 10.1016/j.eswa.2023.120346.

[31] A. O. Alsaiari, E. B. Moustafa, H. Alhumade, H. Abulkhair, and A. Elsheikh, "A coupled artificial neural network with artificial rabbits optimizer for predicting water productivity of different designs of solar stills," Adv. Eng. Softw., vol. 175, p. 103315, Jan. 2023, doi: 10.1016/j.advengsoft.2022.103315.

[32] A. G. Hussien, H. A. El-Sattar, F. A. Hashim, and S. Kamel, "Enhancing optimal sizing of stand-alone hybrid systems with energy storage considering techno-economic criteria based on a modified artificial rabbits optimizer," J. Energy Storage, vol. 78, no. October 2023, p. 109974, Feb. 2024, doi: 10.1016/j.est.2023.109974.

[33] F. Li, H. Xu, and F. Qiu, "Modified artificial rabbits optimization combined with bottlenose dolphin optimizer in feature selection of network intrusion detection," Electron. Res. Arch., vol. 32, no. 3, pp. 1770–1800, 2024, doi: 10.3934/ERA.2024081.

[34] X. Zhou, J. Lu, J. Huang, M. Zhong, and M. Wang, "Enhancing artificial bee colony algorithm with multi-elite guidance," Inf. Sci. (Ny)., vol. 543, pp. 242–258, Jan. 2021, doi: 10.1016/j.ins.2020.07.037.

[35] D. Bajer and B. Zorić, "An effective refined artificial bee colony algorithm for numerical optimisation," Inf. Sci. (Ny)., vol. 504, pp. 221–275, Dec. 2019, doi: 10.1016/j.ins.2019.07.022.

[36] D. Kumar and K. K. Mishra, "Co-variance guided Artificial Bee Colony," Appl. Soft Comput., vol. 70, pp. 86–107, Sep. 2018, doi: 10.1016/j.asoc.2018.04.050.

[37] S. W. Scheff, "Nonparametric Statistics," in Fundamental Statistical Principles for the Neurobiologist, Academic Press, 2016, pp. 157–182. doi: 10.1016/B978-0-12-804753-8.00008-7.

[38] D. V. Tien, P. Hawliczek, R. Gono, and Z. Leonowicz, "Analysis and modeling of STATCOM for regulate the voltage in power systems," in 2017 18th International Scientific Conference on Electric Power Engineering (EPE), IEEE, May 2017, pp. 1–4. doi: 10.1109/EPE.2017.7967358.

[39] P. P. S. Saputra, F. D. Murdianto, R. Firmansyah, and K. Widarsono, "Economic Dispatch in IEEE 26 Bus System using Quantum Behaved Particle Swarm Optimization," in 2020 International Conference on Applied Science and Technology (ICAST), IEEE, 2020, pp. 54–58. doi: 10.1109/iCAST51016.2020.9557625.

[40] M. R. M. Ridzuan, E. E. Hassan, A. R. Abdullah, and A. F. A. Kadir, "Sustainable environmental economic dispatch optimization with hybrid metaheuristic modification," Indones. J. Electr. Eng. Comput. Sci., vol. 11, no. 1, pp. 161–168, 2018, doi: 10.11591/ijeecs.v11.i1.pp161-168.

[41] M. R. Hashim, "Improved spiral dynamics and artificial bee colony algorithms with application to engineering problems," University of Sheffield, 2018.

https://doi.org/10.31436/iiumej.v26i3.3590

CONSTRUCTION AND ANALYSIS OF A NEW SEVEN-DIMENSIONAL CHAOTIC SYSTEM WITH COMPLEX DYNAMICAL BEHAVIORS

SURA HASBALLAH SHNAWA*, SADIQ A. MEHDI, EMAD I ABDUL KAREEM

Department of Computer Science, College of Education, Mustansiriyah University, Baghdad, Iraq

**Corresponding author: suraahassaballah@gmail.com*

(Received: 22 January 2025; Accepted: 3 July 2025; Published online: 9 September 2025)

ABSTRACT: This work presents a novel seven-dimensional hyperchaotic system with six Lyapunov exponents, three of which are positive, confirming chaotic behavior. The proposed system exhibits intricate dynamic behavior. This makes it highly suitable for various applications, particularly in secure communication. The system's key properties were analyzed using Lyapunov exponents, where the highest exponent is "6.68949", further confirming its strong chaotic nature. Additionally, bifurcation diagrams, phase portraits, equilibrium points, the Kaplan-Yorke dimension, and multistability were examined, providing a comprehensive understanding of its complex dynamics. Other important characteristics, including dissipativity, waveform analysis, sensitivity to initial conditions, and keyspace analysis, were also investigated to validate the system's robustness and applicability. A Mathematica model generated phase images of several strange attractors to support the theoretical findings.

ABSTRAK: Karya ini mempersembahkan sistem hiperchaotic tujuh dimensi novel dengan enam eksponen Lyapunov, tiga daripadanya positif, mengesahkan tingkah laku huru-hara. Sistem yang dicadangkan mempamerkan tingkah laku dinamik yang rumit. Ini menjadikannya sangat sesuai untuk pelbagai aplikasi, terutamanya dalam komunikasi selamat. Sifat utama sistem dianalisis menggunakan eksponen Lyapunov, di mana eksponen tertinggi ialah "6.68949", seterusnya mengesahkan sifat huru-hara yang kuat. Selain itu, gambar rajah bifurkasi, potret fasa, titik keseimbangan, dimensi Kaplan-Yorke, dan kestabilan berbilang telah diperiksa, memberikan pemahaman menyeluruh tentang dinamik kompleksnya. Ciri penting lain, termasuk dissipativity, analisis bentuk gelombang, kepekaan kepada keadaan awal dan analisis ruang kekunci, juga disiasat untuk mengesahkan keteguhan dan kebolehgunaan sistem. Menggunakan model MATHEMATICA, imej fasa beberapa penarik pelik dijana untuk menyokong penemuan teori.

KEYWORDS: *HyperChaotic , Seven Dimensional, Dynamic Analysis, Lyapunov exponent, Equilibrium points.*

1. INTRODUCTION

Chaos theory has found numerous practical applications in fields such as engineering, physics, medicine, and encryption [1–3]. Since Lorenz introduced the first 3-dimensional chaotic attractor in 1963 [4], researchers have extensively examined such systems due to their fascinating dynamics and potential uses. Over the decades, various 3D chaotic systems have been developed, including the Lü system, Chen system, Chua's circuit, and Liu system and the list goes on [5]. These systems have inspired the discovery of a wide range of chaotic attractors, such as scrolls, butterflies, and other intricate structures [6]. three-dimensional autonomous

https://doi.org/10.31436/iiumej.v26i3.3590

chaotic systems remain significant due to their simplicity and minimal computational requirements. However, with the growing demand for more reliable and secure applications, the focus has shifted toward higher-dimensional systems, which offer greater complexity and possibilities.

In 1976, Rössler introduced the first hyperchaotic system [7], marking a major breakthrough in chaos theory. Hyperchaotic systems, defined by multiple positive Lyapunov exponents, exhibit more complex dynamics than conventional chaotic systems [8,9]. These systems, typically of order 4 or higher, have gained interest for their potential use in secure communications due to their strong sensitivity to initial conditions and complex behavior [10–12].

The development of hyperchaotic systems has undergone several stages, including theoretical research [13–15], numerical modeling [16,17], experimental validation [18], and application exploration [19,20]. Building on this foundation, this work introduces a novel 7-dimensional autonomous chaotic system, designed to enhance sensitivity and complexity. The proposed system exhibits three positive Lyapunov exponents, providing a high degree of unpredictability and security, making it highly suitable for cryptographic applications.

The study is structured as follows: Section 2 presents the mathematical formulation of the proposed system. Section 3 analyzes its dynamics using properties such as stability, dissipativity, and bifurcation, along with phase portraits and other key characteristics of the proposed hyperchaotic system. Numerical simulations conducted in Mathematica provide insights into its dynamic behavior, laying the groundwork for its application in secure communications and other technological domains.

2. CONSTRUCTION OF THE NOVEL SEVEN-DIMENSIONAL SYSTEM

The proposed autonomous system can be expressed as follows:

$$\begin{aligned}
\frac{dx_1}{dt} &= -ax_1 + bx_2 + cx_4 - dx_3x_5 - x_3x_7 \\
\frac{dx_2}{dt} &= -ex_2 + fx_1 - gx_5 - hx_1x_3 + cx_3x_4 \\
\frac{dx_3}{dt} &= -ix_3 + dx_5 + gx_7 + x_1x_2 - x_2x_4 \\
\frac{dx_4}{dt} &= -x_4 + x_5 + dx_6 - jx_2x_3 - kx_5x_7 \\
\frac{dx_5}{dt} &= -hx_5 + dx_1 - dx_3 + x_2x_7 + x_6x_7 \\
\frac{dx_6}{dt} &= -lx_6 - x_2 + x_5 + mx_1x_7 - nx_3x_4 \\
\frac{dx_7}{dt} &= -ex_7 + jx_3 - dx_5 + hx_1x_2 + x_4x_5
\end{aligned} \tag{1}$$

The system's positive parameters are denoted as $a, b, c, d, e, f, g, h, I, j, k, l, m,$ and n, while its state variables are given by $x1, x2, x3, x4, x5, x6,$ and $x7$. When the following system parameter values are selected, the seven-dimensional system (1) exhibits a strange attractor:

$a = 15,\ b = 13,\ c = 0.4,\ d = 0.5,\ e = 14,\ f = 38,\ g = 1.5,\ h = 2,\ I = 6,\ j = 3,\ k = 5,\ l = 15.1,\ m = 30,$ and $n = 4$. The initial values are given as follows: $x1(0) = 5, x2(0) = 0.7, x3(0) = 3, x4(0) = 10, x5(0) = 0.2, x6(0) = 0.1,$ and $x7(0) = 0.6$.

https://doi.org/10.31436/iiumej.v26i3.3590

3. ANALYSIS OF THE NEW HYPERCHAOTIC SYSTEM

This section presents an examination of the novel hyperchaotic system, describing its qualitative features explored in this work.

3.1. System Equilibrium

Equilibrium points of the proposed hyperchaotic system (1) are derived through the solution of nonlinear equations:

$$\begin{aligned}
0 &= -ax_1 + bx_2 + cx_4 - dx_3x_5 - x_3x_7 \\
0 &= -ex_2 + fx_1 - gx_5 - hx_1x_3 + cx_3x_4 \\
0 &= -ix_3 + dx_5 + gx_7 + x_1x_2 - x_2x_4 \\
0 &= -x_4 + x_5 + dx_6 - jx_2x_3 - kx_5x_7 \\
0 &= -hx_5 + dx_1 - dx_3 + x_2x_7 + x_6x_7 \\
0 &= -lx_6 - x_2 + x_5 + mx_1x_7 - nx_3x_4 \\
0 &= -ex_7 + jx_3 - dx_5 + hx_1x_2 + x_4x_5
\end{aligned} \tag{2}$$

The single points of equilibrium turn into:

$E0 = \{x1 = 0, x2 = 0, x3 = 0, x4 = 0, x5 = 0, x6 = 0, x7 = 0\}$. Jacobian matrix of system(1),let

$$f = \begin{cases}
f1 = \frac{dx_1}{dt} = -ax_1 + bx_2 + cx_4 - dx_3x_5 - x_3x_7 \\
f2 = \frac{dx_2}{dt} = -ex_2 + fx_1 - gx_5 - hx_1x_3 + cx_3x_4 \\
f3 = \frac{dx_3}{dt} = -ix_3 + dx_5 + gx_7 + x_1x_2 - x_2x_4 \\
f4 = \frac{dx_4}{dt} = -x_4 + x_5 + dx_6 - jx_2x_3 - kx_5x_7 \\
f5 = \frac{dx_5}{dt} = -hx_5 + dx_1 - dx_3 + x_2x_7 + x_6x_7 \\
f6 = \frac{dx_6}{dt} = -lx_6 - x_2 + x_5 + mx_1x_7 - nx_3x_4 \\
f7 = \frac{dx_7}{dt} = -ex_7 + jx_3 - dx_5 + hx_1x_2 + x_4x_5
\end{cases} \tag{3}$$

$$J = \begin{bmatrix}
\frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} & \frac{\partial f_1}{\partial x_4} & \frac{\partial f_1}{\partial x_5} & \frac{\partial f_1}{\partial x_6} & \frac{\partial f_1}{\partial x_7} \\
\frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} & \frac{\partial f_2}{\partial x_4} & \frac{\partial f_2}{\partial x_5} & \frac{\partial f_2}{\partial x_6} & \frac{\partial f_2}{\partial x_7} \\
\frac{\partial f_3}{\partial x_1} & \frac{\partial f_3}{\partial x_2} & \frac{\partial f_3}{\partial x_3} & \frac{\partial f_3}{\partial x_4} & \frac{\partial f_3}{\partial x_5} & \frac{\partial f_3}{\partial x_6} & \frac{\partial f_3}{\partial x_7} \\
\frac{\partial f_4}{\partial x_1} & \frac{\partial f_4}{\partial x_2} & \frac{\partial f_4}{\partial x_3} & \frac{\partial f_4}{\partial x_4} & \frac{\partial f_4}{\partial x_5} & \frac{\partial f_4}{\partial x_6} & \frac{\partial f_4}{\partial x_7} \\
\frac{\partial f_5}{\partial x_1} & \frac{\partial f_5}{\partial x_2} & \frac{\partial f_5}{\partial x_3} & \frac{\partial f_5}{\partial x_4} & \frac{\partial f_5}{\partial x_5} & \frac{\partial f_5}{\partial x_6} & \frac{\partial f_5}{\partial x_7} \\
\frac{\partial f_6}{\partial x_1} & \frac{\partial f_6}{\partial x_2} & \frac{\partial f_6}{\partial x_3} & \frac{\partial f_6}{\partial x_4} & \frac{\partial f_6}{\partial x_5} & \frac{\partial f_6}{\partial x_6} & \frac{\partial f_6}{\partial x_7} \\
\frac{\partial f_7}{\partial x_1} & \frac{\partial f_7}{\partial x_2} & \frac{\partial f_7}{\partial x_3} & \frac{\partial f_7}{\partial x_4} & \frac{\partial f_7}{\partial x_5} & \frac{\partial f_7}{\partial x_6} & \frac{\partial f_7}{\partial x_7}
\end{bmatrix} \tag{4}$$

https://doi.org/10.31436/iiumej.v26i3.3590

$$J = \begin{bmatrix} -a & b & -dx_5 - x_7 & c & -dx_3 & 0 & -x_3 \\ f - hx_3 & -e & -hx_1 - x_4 & -cx_3 & -g & 0 & 0 \\ x_2 & x_1 - x_4 & -i & -x_2 & d & 0 & g \\ 0 & -jx_3 & -jx_2 & -1 & 1 - kx_7 & d & -kx_5 \\ d & x_7 & -d & 0 & -h & x_7 & x_2 + x_6 \\ mx_7 & -1 & -nx_4 & -nx_3 & 1 & -l & mx_7 \\ hx_2 & hx_1 & j & x_5 & -d + x_4 & 0 & -e \end{bmatrix} \quad (5)$$

For the point of equilibrium:

$E0 = \{x1 = 0, x2 = 0, x3 = 0, x4 = 0, x5 = 0, x6 = 0, x7 = 0\}$, with values $a = 15, b = 13, c = 0.4, d = 0.5, e = 14, f = 38, g = 1.5, h = 2, I = 6, j = 3, k = 5, l = 15.1, m = 30,$ and $n = 4$.

The following is the outcome of the Jacobian matrix:

$$J_0 = \begin{bmatrix} -15 & 13 & 0 & 0.4 & 0 & 0 & 0 \\ 38 & -14 & 0 & 0 & -1.5 & 0 & 0 \\ 0 & 0 & -6 & 0 & 0.5 & 0 & 1.5 \\ 0 & 0 & 0 & -1 & 1 & 0.5 & 0 \\ 0.5 & 0 & -0.5 & 0 & -2 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & -15.1 & 0 \\ 0 & 0 & 3 & 0 & -0.5 & 0 & -14 \end{bmatrix} \quad (6)$$

To obtain its eigenvalues, we set $|\lambda I - J0| = 0$. The resulting eigenvalues equivalent with the stable point E0 (0, 0, 0, 0, 0, 0, 0), are calculated as shown below:

$\lambda_1 = -36.7377, \quad \lambda_2 = -15.011, \quad \lambda_3 = -14.5232,$

$\lambda_4 = -5.41581, \quad \lambda_5 = -2.02482, \quad \lambda_6 = -1.00679,$

$\lambda_7 = 7.7095$

The results show that the system (1), at the equilibrium point E0, has both positive and negative real eigenvalues. Therefore, this equilibrium point is unstable.

3.2. Dissipativity

The novel system (1) can be expressed in vector notation as:

$$f = \begin{cases} f_1 = \frac{dx_1}{dt} = -ax_1 + bx_2 + cx_4 - dx_3x_5 - x_3x_7 \\ f_2 = \frac{dx_2}{dt} = -ex_2 + fx_1 - gx_5 - hx_1x_3 + cx_3x_4 \\ f_3 = \frac{dx_3}{dt} = -ix_3 + dx_5 + gx_7 + x_1x_2 - x_2x_4 \\ f_4 = \frac{dx_4}{dt} = -x_4 + x_5 + dx_6 - jx_2x_3 - kx_5x_7 \\ f_5 = \frac{dx_5}{dt} = -hx_5 + dx_1 - dx_3 + x_2x_7 + x_6x_7 \\ f_6 = \frac{dx_6}{dt} = -lx_6 - x_2 + x_5 + mx_1x_7 - nx_3x_4 \\ f_7 = \frac{dx_7}{dt} = -ex_7 + jx_3 - dx_5 + hx_1x_2 + x_4x_5 \end{cases} \quad (7)$$

The vector field f's divergence in $\mathbf{R}^7$ is written as:

https://doi.org/10.31436/iiumej.v26i3.3590

$$\nabla \cdot \mathrm{f} = \frac{\partial \mathrm{f}_1}{\partial \mathrm{x}_1} + \frac{\partial \mathrm{f}_2}{\partial \mathrm{x}_2} + \frac{\partial \mathrm{f}_3}{\partial \mathrm{x}_3} + \frac{\partial \mathrm{f}_4}{\partial \mathrm{x}_4} + \frac{\partial \mathrm{f}_5}{\partial \mathrm{x}_5} + \frac{\partial \mathrm{f}_6}{\partial \mathrm{x}_6} + \frac{\partial \mathrm{f}_7}{\partial \mathrm{x}_7} \quad (8)$$

It should be noted that the rate at which volumes change under the flow Φ_t generated by f is measured by $\nabla \cdot f$. let $\boldsymbol{D}$ be a region in $\mathbf{R}^7$ with a smooth boundary. Next, examine the image of $\boldsymbol{D}$ corresponding to Φ_t, the time t of the flow of f, where $\boldsymbol{D}(\boldsymbol{t}) = \Phi_t(t)$. In this case, $V(t)$ represents the volume of $D(t)$. According to Liouville's theorem [55,67], we obtain the following:

$$\frac{dV}{dt} = \int_{D(t)} (\nabla \cdot f) dx_1 dx_2 dx_3 dx_4 dx_5 dx_6 dx_7 (9)$$

For the system (1), we find that:

$$\nabla \cdot f = -(a + e + i + 1 + h + l + e) < 0 \quad (10)$$

Because a, e, i, h and l are positive constants, substituting Equation (10) into Equation (9) and simplifying, we get:

$$\frac{dV}{dt} = -(a + e + i + 1 + h + l + e) \int_{D(t)} dx_1 dx_2 dx_3 dx_4 dx_5 dx_6 \quad (11)$$

$$\frac{dV}{dt} = -(a + e + i + 1 + h + l + e) V(t)$$

$$= e^{-65.1t} V(t)$$

The unique solution to the first-order linear differential equation (11) is derived as follows:

$$\mathrm{V(t)} = \mathrm{V(0)e^{-(a+e+i+1+h+l+e)t}}$$

$$= \mathrm{V(0)e^{-65.1t}} \quad (12)$$

According to Equation (12), any volume V(t) decreases exponentially to zero over time. Consequently, equations (1) describe a dissipative dynamical system.

3.3. Lyapunov Dimensions and Exponents

A chaotic system relies on sensitivity to initial conditions, which the Lyapunov Exponents measure in accordance with nonlinear dynamical theory. A Lyapunov exponent represents the average rate at which two nearby trajectories diverge (or converge) over time. This sensitivity highlights the unpredictable nature of chaotic systems and emphasizes their complex dynamical behavior.

Furthermore, seven Lyapunov exponents of the nonlinear dynamical system (1), with specified parameters, were calculated as follows:

$$LE1 = 6.68949, LE2 = 2.65441, LE3 = 1.4784, LE4 = -8.79264, LE5 = -12.8942,$$

$$LE6 = -14.8192, LE7 = -39.706$$

As shown, this system exhibits chaotic behavior because its largest Lyapunov exponent is positive. This indicates that trajectories in the phase space diverge exponentially, a hallmark of chaos. Since the remaining four Lyapunov exponents are negative, the system shows a stable contraction along those dimensions. Additionally, *LE_1, LE_2, and LE_3* are positive Lyapunov exponents, highlighting the system's complexity. The system is therefore classified as hyperchaotic due to having more than one positive Lyapunov exponent.

Another common feature of chaotic systems is the fractal dimension, which provides insight into their geometric complexity. The fractal dimension is derived from the Lyapunov exponents

https://doi.org/10.31436/iiumej.v26i3.3590

and is used to calculate the Kaplan–Yorke dimension. The Kaplan-Yorke dimension, denoted as D_{KY}, is given by [15]:

$$D_{KY} = j + \frac{1}{|L_{Ej+1}|}\sum_{i=1}^{j} LE_i \tag{13}$$

Here, the first j Lyapunov exponents are non-negative, as indicated by j, which is the greatest value of i that simultaneously satisfies $\sum_{i=1}^{j} LE_i > 0$ and $\sum_{i=1}^{j+1} LE_i < 0$. Based on the Lyapunov exponents sequence, LE_i is in declining order of the sequence. The upper limit of the system information's dimension is denoted by D_{KY}.

Since $LE_1 + LE_2 + LE_3 + LE_4 > 0$ *and* $LE_1 + LE_2 + LE_3 + LE_4 + LE_5 + LE_6 + LE_7 < 0$, The Kaplan-Yorke dimension for the system in this work can be obtained by analyzing the seven calculated Lyapunov exponents presented above, which show that j equals four. Concerning the new chaotic system, the Lyapunov dimension is:

$$\begin{aligned} D_{KY} &= j + \frac{1}{|LE_{j+1}|}\sum_{i=1}^{j} LE_i \\ D_{KY} &= 4 + \frac{1}{|LE_{j+1}|}\sum_{i=1}^{4} LE_i = 5 + \frac{LE_1+LE_2+LE_3+LE_4}{|LE_5|} \\ &= 4 + \frac{(6.68949+2.65441+1.4784-8.79264)}{12.8942} \\ &= 4.15741 \end{aligned}$$

This result suggests that the Lyapunov dimension of the system (1) is fractional. The system exhibits aperiodic trajectories due to its fractal nature, and its neighboring trajectories diverge as well. Therefore, this nonlinear system can be considered genuinely chaotic.

3.4. Phase Portraits

Using the *MATHEMATICA* program, the numerical simulation has been completed to analyze the chaotic behavior of this nonlinear system. This system demonstrates intricate and abundant chaotic dynamics, which strange attractors characterize. These attractors are visualized in three-dimensional space in Figures (1–3), and in two-dimensional projections in Figures (4–6). The system's dynamics reveal complex patterns, with the attractors exhibiting irregular yet structured behavior, typical of chaotic systems. The topology observed in the phase portraits resembles the so-called "butterfly effect." As depicted in Figure 1, this structure resembles the contours of a butterfly in flight, illustrating the sensitivity of chaotic systems to initial conditions. The "butterfly effect" emphasizes how small changes in initial conditions can lead to vastly different outcomes—a defining characteristic of chaotic systems.

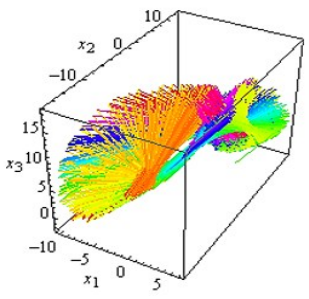


Figure 1. 3D Representation of Chaotic Attractors (x_1-x_2-x_3)

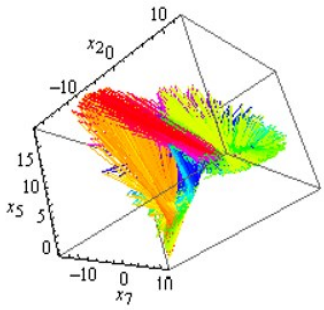


Figure 2. 3D Representation of Chaotic Attractors (x_2-x_5-x_7)

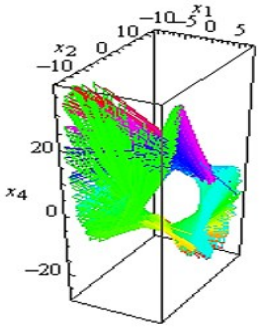


Figure 3. 3D Representation of Chaotic Attractors (x_1-x_2-x_4)

https://doi.org/10.31436/iiumej.v26i3.3590

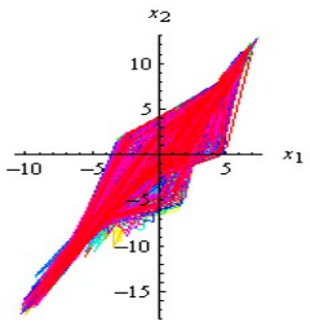


Figure 4. 2D Representation of Chaotic Attractors (x_2- x_1)

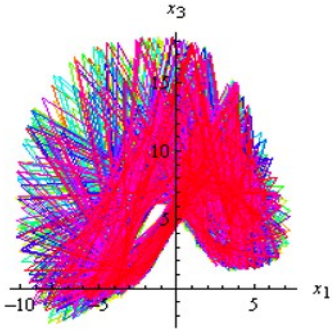


Figure 5. 2D Representation of Chaotic Attractors (x_3 -x_1)

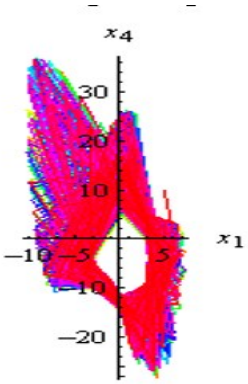


Figure 6. 2D Representation of Chaotic Attractors (x_4- x_1)

3.5. Waveform Analysis in a New Chaotic System

It is well known that a chaotic system typically exhibits aperiodic waveforms. To illustrate the chaotic nature of the suggested system. time-versus-state plots obtained from *MATHEMATICA* simulations are shown in Figures 7–13. Figures (7–13) display the time-domain waveforms of $x_1(t), x_2(t), x_3(t), x_4(t), x_5(t), x_6(t), x_7(t)$. Each of these waveforms is aperiodic, confirming the system's chaotic nature. It is possible to distinguish between chaotic motion and multiple periodic motions, which can both exhibit complex behavior, by looking for non-cyclical features in the time domain waveform.

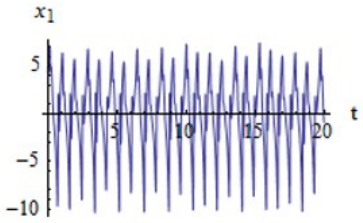


Figure 7. Time vs. x_1 in a novel chaotic system

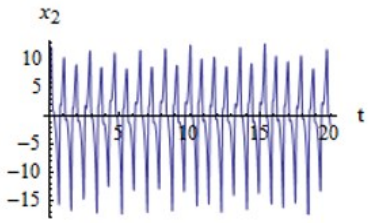


Figure 8. Time vs. x_2 in a novel chaotic system

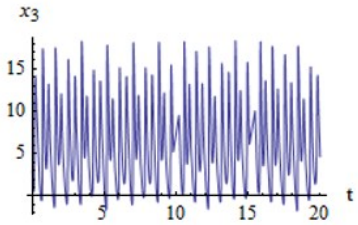


Figure 9. Time vs. x_3 in a novel chaotic system

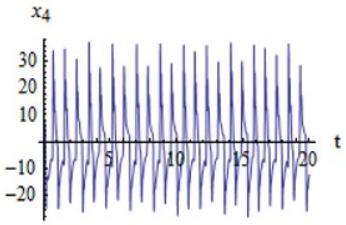


Figure 10. Time vs. x_4 in a novel chaotic system

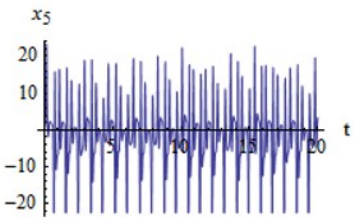


Figure 11. Time vs. x_5 in a novel chaotic system

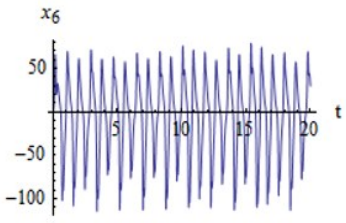


Figure 12. Time vs. x_6 in a novel chaotic system

https://doi.org/10.31436/iiumej.v26i3.3590

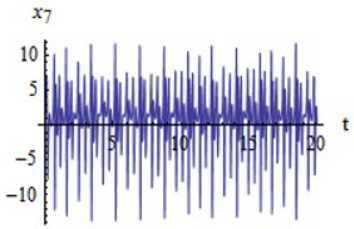


Figure 13. Time vs. x_7 in a novel chaotic system

3.6. Initial condition sensitivity

The long-term unpredictability of a chaotic system is arguably its most defining quality. This intrinsic property arises from the system's solutions being highly sensitive to initial conditions. Even the slightest differences between two starting conditions — regardless of how similar they may appear—will eventually cause the trajectories to become vastly separated over time. This sensitivity leads to a situation where, in the future, it becomes increasingly complex to make accurate predictions about the system's state. Any attempt to predict the system's behavior is constrained by the precision of the initial conditions. As shown in Figures 16–22, even slight variations in initial conditions result in dramatically different trajectories. These figures demonstrate the significant impact that initial conditions have on the system's chaotic behavior. The initial conditions used in the simulations are as follows:

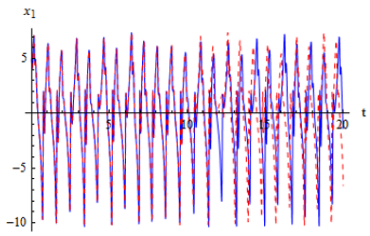


Figure 14 $x_1(t)$ Sensitivity Analysis of the Novel System

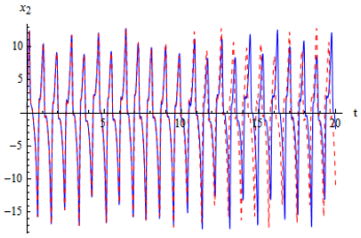


Figure 15. $x_2(t)$ Sensitivity Analysis of the Novel System

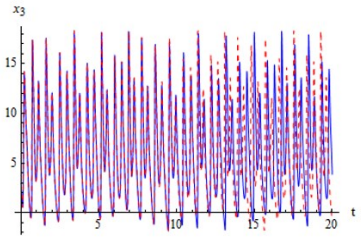


Figure 16. $x_3(t)$ Sensitivity Analysis of the Novel System

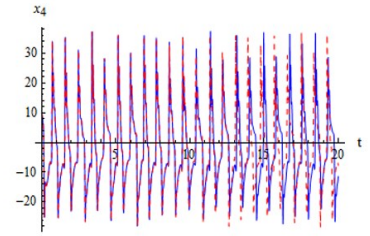


Figure 17. $x_4(t)$ Sensitivity Analysis of the Novel System

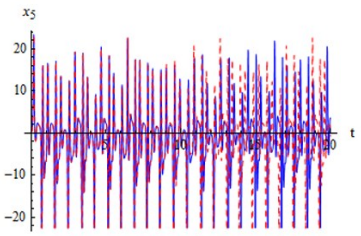


Figure 18. $x_5(t)$ Sensitivity Analysis of the Novel System

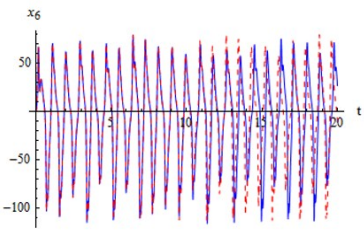


Figure 19. $x_6(t)$ Sensitivity Analysis of the Novel System

https://doi.org/10.31436/iiumej.v26i3.3590

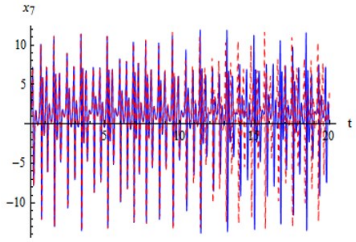


Figure 20. $x_7(t)$ Sensitivity Analysis of the Novel System

3.7. Bifurcation Diagram

The *MATHEMATICA* software was used to simulate the new system (1) and determine the maximum value of the variable x_1. By varying the values of the parameter a, we observe the intriguing region of x_1 where the system changes its dynamics and values of x_1 begin to bifurcate. This bifurcation phenomenon is one of the most prominent characteristics of chaotic systems and clearly indicates the system's sensitivity to parameter variation. This behavior is particularly evident in the range $a = 14.2$ and $a = 14.4$, where the system undergoes a transition to more complex dynamics. The bifurcation diagram provides insight into how the system evolves and displays its chaotic dynamics. Figure 23 illustrates how bifurcation occurs in the new system, showcasing the dramatic changes in its properties as the parameter a is adjusted.

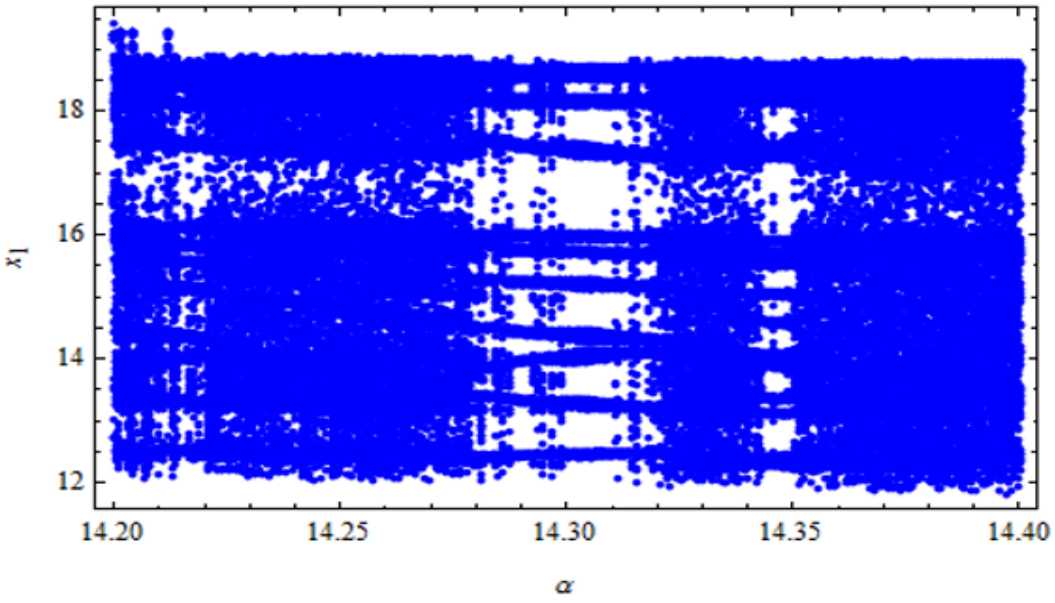


Figure 21. Bifurcation diagram of the variable x_1 for increasing $\boldsymbol{a}$

3.8. Analysis of Key Spaces

The total number of distinct keys that can be used throughout the encryption process is referred to as the *keyspace* size. A strong encryption algorithm should be sensitive to secret keys.

The *keyspace* should be large enough to prevent attacks using brute force. Generally speaking, a smaller *keyspace* makes the system more vulnerable to attacks. Cryptographically, to make brute force attacks ineffective, the size of the *keyspace* should be at least 2^{128} Since

the parameters $a, b, c, d, e, f, g, h, i, j, k, l, m,$ and n and initial conditions $x_1(0), x_2(0), x_3(0), x_4(0), x_5(0), x_6(0)$, and $x_7(0)$ serve as the hidden keys in this algorithm and used as secret key. In this instance, the *keyspace* size can reach $(10^{14})^{21} = 10^{294} \simeq 2^{970}$, assuming a precision of 10^{-14}, which is significantly larger than 2^{128}. Therefore, the *keyspace* is sufficiently large to withstand the brute-force attacks.

4. CONCLUSION

This paper introduces a new seven-dimensional (7D) chaotic system. The system exhibits intricate behavior and high sensitivity to initial conditions, showcasing its potential for advanced applications in the field of chaos theory. Through thorough analyses, including the Kaplan-Yorke dimension and equilibrium point evaluations, we demonstrate the system's strongly chaotic nature. This is reflected in its three positive Lyapunov exponents, which confirm its hyperchaotic nature and its ability to generate a wide variety of strange attractors.

The system's dissipative nature and inherent instability further highlight its suitability for critical applications in secure communication and encryption. In such contexts, sensitivity to initial conditions and dynamic complexity is an essential characteristic. These attributes enable the system to effectively obscure data, making it robust against various types of cryptographic attacks, including brute-force attacks, as demonstrated by the findings of this study.

Additionally, numerical simulations conducted using Mathematica provided a detailed exploration of the system's dynamics, including its phase diagrams and chaotic trajectories. These simulations have been instrumental in validating the theoretical properties of the system and in illustrating its potential for real-world applications.

Overall, the proposed seven-dimensional framework marks a significant advancement in the study of chaotic systems and their applications. It provides a solid foundation for future research, particularly in the development of innovative encryption algorithms and emerging security technologies. Subsequent investigations will focus on its broader applicability and further optimization for real-world cryptographic systems, ensuring its viability in the evolving landscape of secure communication.

REFERENCES

[1] Skinner JE, Molnar M, Vybiral T, Mitra M. (1992) Application of chaos theory to biology and medicine. Integrative Physiological and Behavioral Science, 27:39-53.

[2] Ewees AA, Abd Elaziz M. (2020) Performance analysis of chaotic multi-verse harris hawks optimization: a case study on solving engineering problems. Engineering Applications of Artificial Intelligence, 88:103370. https://doi.org/10.1016/j.engappai.2019.103370

[3] AL-Hashemy RH, Mehdi SA. (2019) A new algorithm based on magic square and a novel chaotic system for image encryption. Journal of Intelligent Systems, 29(1):1202-15. https://doi.org/10.1515/jisys-2018-0404

[4] Lorenz EN. (1963) Deterministic nonperiodic flow. Journal of atmospheric sciences,20(2):130-41. https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2

[5] Lu J, Chen G, Yu X, Leung H. (2004) Design and analysis of multiscroll chaotic attractors from saturated function series. IEEE Transactions on Circuits and Systems I: Regular Papers, 51(12):2476-90. https://doi.org/10.1109/TCSI.2004.838151

[6] Shakir HR, Hattab AA, Mehdi SA. (2022) A dynamic S-box generation based on a hybrid method of new chaotic system and DNA computing. TELKOMNIKA (Telecommunication Computing Electronics and Control), 20(6):1230-8. http://doi.org/10.12928/telkomnika.v20i6.23449

[7] Rössler OE. (1976) An equation for continuous chaos. Physics Letters A, 57(5):397-8. https://doi.org/10.1016/0375-9601(76)90101-8

[8] Jia Q. (2007) Hyperchaos generated from the Lorenz chaotic system and its control. Physics Letters A, 366(3):217-22. https://doi.org/10.1016/j.physleta.2007.02.024

[9] Lagmiri SN, Amghar M, Sbiti N. (2017) Hyperchaos based cryptography: New seven-dimensional systems to secure Communications. Circ. Comput Sci, 2(2):20-30.

[10] Sadkhan SB, Mehdi SA, Zghair HK. (2020) Analysis of Novel Seven-Dimension Hyper Chaotic by Using SDIC and Waveform. In proceedings of the 3rd International Conference on Engineering Technology and its Applications (IICETA): 2020; IEEE, pp. 95-99. https://doi.org/10.1109/IICETA50496.2020.9318940

[11] Sun S, Guo Y, Wu R. (2019) A novel image encryption scheme based on 7D hyperchaotic system and row-column simultaneous swapping. IEEE Access, 7:28539-47. https://doi.org/10.1109/ACCESS.2019.2901870

[12] Kopp M. (2024) Dynamical analysis, circuit design, synchronization, and Arduino UNO implementation of a new 7D hyperchaotic system. International Journal of Chaos Theory, 10(3): 123-134.

[13] Singh JP, Rajagopal K, Roy BK. (2018) A new 5D hyperchaotic system with stable equilibrium point, transient chaotic behaviour and its fractional-order form. Pramana, 91:1-0.

[14] Mehdi SA, Zghair HK, Sadkhan SB. (2021) Bifurcation of novel seven-dimension hyper chaotic system. Journal of Physics: Conference Series, 1804: 012051. IOP Publishing.

[15] Yu W, Wang J, Wang J, Zhu H, Li M, Li Y, Jiang D. (2019) Design of a new seven-dimensional hyperchaotic circuit and its application in secure communication. IEEE Access, 7:125586-608. https://doi.org/10.1109/ACCESS.2019.2935751

[16] Thivagar ML, Hamad AA, Tamilarasan B, Antony GK. (2022) A novel seven-dimensional hyperchaotic. InProceedings of Second Doctoral Symposium on Computational Intelligence: DoSCI 2021 2022, pp. 329-340. Springer Singapore.

[17] Zghair HK, Mehdi SA, Sadkhan SB. (2020) Design and analytic of a novel seven-dimension hyper chaotic systems. In2020 1st Information Technology to Enhance e-learning and Other Application (IT-ELA 2020), pp. 77-81. IEEE. https://doi.org/10.1109/IT-ELA50150.2020.9253077

[18] Vaidyanathan S, He S, Simbas A. (2021) A new multistable double-scroll 4-D hyperchaotic system with no equilibrium point, its bifurcation analysis, synchronization and circuit design. Archives of Control Sciences, 31. http://dx.doi.org/10.24425/acs.2021.136882

[19] Wu S, Li G, Xu W, Xu X, Zhong H. (2023) Modelling and dynamic analysis of a novel seven-dimensional Hamilton conservative hyperchaotic systems with wide range of parameter. Physica Scripta, 98(5):055218.

[20] Mehdi SA, Mohammed SJ. (2020) Web application authentication using ZKP and novel 6D chaotic system. Indonesian Journal of Electrical Engineering and Computer Science, 20(3):1522-9.

https://doi.org/10.31436/iiumej.v26i3.3253

SAFETY ENHANCEMENT OF PORTABLE OIL SPILL SKIMMER (POSS) VIA COMPUTATIONAL FLUID DYNAMICS FOR LIQUID SLOSHING ANALYSIS

RPRAKASH RAMANATHAN [1,2], LOKMAN ABDULLAH[2*], MUHAMMAD HAFIDZ FAZLI MD FAUADI[2], MUHAMMAD SYAFIQ SYED MOHAMED[2], RUZAIDI ZAMRI[2], RAHIMAH ABDUL HAMID[2], ZAMBERI JAMALUDIN[2], MOHD SHAHRIEEL MOHD ARAS[3]

[1]Seatrium Limited, 80 Tuas South Boulevard, 637051, Singapore
[2]Fakulti Teknologi dan Kejuruteraan Industri dan Pembuatan, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
[3]Fakulti Teknologi dan Kejuruteraan Elektrik, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia

**Corresponding author: lokman@utem.edu.my*

(Received: 25 August 2024; Accepted: 31 May 2025; Published online: 9 September 2025)

ABSTRACT: The Portable Oil Spill Skimmer (POSS) was designed and developed due to several disadvantages of the current methods of Oil Spill Response and Recovery (OSRR). The POSS was designed as a complementary method to aid the OSRR tasks. However, during the POSS maneuverability testing, the POSS experiences instability when moving in different directions. The imbalance occurs when there is the presence of oil in the oil tank. Based on the literature study, the liquid sloshing effect was the reason why the POSS experiences instability. Thus, this research aims to analyse the impact of liquid sloshing in an oil tank and the implementation of baffles to reduce the effect. The analysis was conducted using SolidWorks Flow Simulation to simulate the liquid sloshing in the oil tank. The simulation was conducted in two situations, with and without baffles, to compare the results. According to the obtained results, with the implementation of 3 baffles, the sloshing effect was reduced to 392 N of torque force from 1195.43 N without baffles. The reduction was significant as the sloshing effect cannot be eliminated, thus the torque force of 392 N was enough to minimise the stability issue of the POSS.

ABSTRAK: Penapis Tumpahan Minyak Mudah Alih (POSS) telah direka bentuk dan difabrikasi kerana terdapat beberapa kelemahan kaedah semasa iaitu Tindak Balas dan Pemulihan Tumpahan Minyak (OSRR). POSS direka bentuk sebagai kaedah pelengkap bagi membantu operasi OSRR. Walau bagaimanapun, semasa ujian kebolehgerakan POSS, ia mengalami ketidakstabilan gerakan arah berbeza. Ketidakseimbangan ini berlaku apabila terdapat minyak dalam tangki minyak. Berdasarkan kajian, POSS mengalami ketidakstabilan disebabkan oleh kesan percikan cecair (liquid sloshing). Oleh itu, kajian ini bertujuan bagi menganalisis kesan percikan minyak (oil sloshing) dalam tangki minyak dan mengkaji keberkesanan pelaksanaan penyekat dalaman (baffles) bagi mengurangkan kesan percikan. Analisis dijalankan dengan menggunakan Perisian Simulasi Aliran SolidWorks bagi mensimulasikan percikan cecair dalam tangki minyak. Simulasi dijalankan dalam dua keadaan, dengan dan tanpa penyekat dalaman bagi membandingkan keputusan. Dapatan kajian mendapati melalui pelaksanaan 3 penyekat dalaman (baffles), kesan percikan telah berjaya dikurangkan kepada 392 N daya kilas (torque force) berbanding 1195.43 N tanpa menggunakan penyekat dalaman. Pengurangan ini adalah ketara kerana kesan percikan tidak dapat dihapuskan sepenuhnya. Oleh itu, daya kilas 392 N adalah cukup bagi meminimumkan isu kestabilan POSS.

https://doi.org/10.31436/iiumej.v26i3.3253

KEYWORDS: *Liquid Sloshing, Portable Oil Spill Skimmer, Oil Spill Response and Recovery, Baffles.*

1. INTRODUCTION

Oil spills have a range of immediate and long-term effects on the environment and wildlife. Environmental damage is a significant consequence, with oil spills causing harm to marine ecosystems, coastal areas, coral reefs, and wetlands. The causes of oil spills are diverse, including accidents during transportation, offshore drilling and production, refinery and storage accidents, and natural disasters. However, efforts are being made to prevent and respond to oil spills, including improved safety regulations, emergency response plans, and the development of technologies for containment and cleanup. Several technologies were invented to address this problem, such as mechanical techniques, chemical techniques, biodegradation, and open burning. Out of all these techniques, the mechanical technique is the most favoured and most effective as it does not alter the oil's physical properties. The mechanical method consists of booms and skimmers. However, this method has the disadvantage of containment issues, and skimmers are static or moved by external force. Thus, a portable oil spill skimmer was developed to complement the current Oil Spill Response and Recovery (OSRR) task.

The developed portable oil spill skimmer (POSS) has a self-propelling mechanism. The POSS could travel to limited-access areas, equipped with an oil tank to store the recovered oil. Meanwhile, existing oil spill skimmers are typically designed without an onboard oil tank, as the recovered oil is pumped into external storage. In addition, a limited study indicates the self-propelling features of the oil spill skimmer. Moreover, the developed POSS has an issue of instability during the maneuvering process. This instability of the POSS occurs when the oil tank is filled with the recovered oil. Thus, this study aims to analyse POSS's self-propelling capability and instability issues. The traveling speed of the POSS will be analysed both in simulation and experiment. Performing mechanical analysis is essential, as the findings are needed for various design processes, including electrical and electronics integration designs, controller designs, aerodynamic analysis, and industrial machine designs [1]–[3]. Computational Fluid Dynamics (CFD) is a numerical analytical approach used in mechanical engineering to approximate solutions to diverse engineering problems. For the simulation, a CFD analysis will be conducted to identify the drag and thrust forces produced. The travel distance of POSS will be used to perform the experimental analysis over time. In addition, the causes of POSS instability were investigated to enhance the fabrication process of POSS and address the issue. The designed POSS is equipped with a pair of propellers to maneuver it. Self-propelling propellers are selected as the propulsion system's propellers. USVs generally use a catamaran-like twin hull configuration for better roll stability compared to single hull designs [4]. Thus, the POSS was fabricated with a twin hull design to increase its stability.

However, during the experimental testing, the POSS still experiences instability when accelerating, decelerating, and turning. Also, the effect of sloshing impacts the structural integrity of the oil tank. From the testing, it is observed that the instability occurs when oil is in the oil tank. The instability of the POSS increases when the oil tank is partially filled. Thus, based on several literature reviews, it is found that the POSS undergoes an effect called liquid sloshing. Liquid sloshing refers to the movement of a free liquid surface within a partially filled tank due to external forces [5]. The research on the impact of liquid sloshing involves various engineering domains, such as aerospace, marine, oil tankers, and more [6]-[8]. According to research, the free oil liquid surface experiences different types of motion depending on the tank shape and type of disturbance [9]. The study found that higher excitation frequencies and

https://doi.org/10.31436/iiumej.v26i3.3253

amplitudes result in more intense sloshing, noticeable splashing waves, and significant breaking of the liquid surface [10].

Furthermore, research was conducted to analyse the dynamic pressure of the sloshing effect with various filling and motion conditions [11]. The result indicates that the partial filling of the oil tank has the most significant dynamic pressure that affects the safety of the fuel tank. Other researchers are also concerned about the safety of the fuel tank; thus, the mitigation of sloshing motion must be achieved [6], [12]-[15]. Several researchers have analysed the importance of implementing baffles to reduce the oil sloshing effect. Based on the research conducted, the incorporation of baffles in the oil tank shows significant improvement, reducing the impact of oil sloshing [16]-[21]. Baffles are usually installed in oil tank conFigureurations to mitigate or diminish sloshing and support the structures of the oil tanks [22-23].

Although the analysis of liquid sloshing has been conducted in various applications, to the best of the author's knowledge, it is less studied in the context of an oil spill skimmer. As the typical oil spill skimmer was developed without an oil tank, this analysis was not explored. Thus, in this research, the significant gaps in understanding the impact and ways to overcome it of oil sloshing are investigated to enhance the safety features of the POSS.

2. METHODOLOGY

2.1. Research Framework

Computational Fluid Dynamics (CFD) was conducted as a preliminary test to analyse the drag force exerted on the hull of the POSS [23]. Ansys Fluent was utilised to perform the analysis. Furthermore, the capability of the propellers was analysed to identify the thrust force produced. Thus, the result of the drag and thrust forces can be compared. The maneuverability of the POSS was experimented with, and the speed traveled was tabulated. Finally, the oil sloshing effect was also analysed by utilising CFD and investigating whether the implementation of baffles reduces the sloshing effect in the designed oil tank of a POSS.

2.2. Computational Fluid Dynamics (CFD) Analysis

2.2.1. CFD Analysis of Hull

ANSYS Fluent is a cutting-edge computer tool that allows you to model fluid flow in complicated shapes. It enables the user to enter several models for mesh creation of various forms and refine or coarsen the mesh of the model based on the flow solution. Figure 1 shows the steps of simulating the hull's resistance. This research utilizes SolidWorks software to create a 3D model, which is then simulated in ANSYS Fluent to calculate the resistance when traveling over the water surface. Several steps were conducted in this research, as follows:

- Step 1: Draw a 3D model of the hole in SolidWorks.

  The hull's dimensions are measured using a tape measure and a vernier caliper. Next, the camera captured the top, side, and front views of the peculiar form of the hall at the gun. Thus, the solid modeling of the hull was done by following the previously measured dimensions.

- Step 2: Export to the NSYS program

  The 3D model of the file is imported to ANSYS Design Modeler in the SAT format.

- Step 3: Generate domains for hull

https://doi.org/10.31436/iiumej.v26i3.3253

The domain is traced around the hull, with the stern hall at the start of the plane. There are two types of domains: a global domain and a subdomain. The global domain ship is a cuboid with dimensions of 4 metres in length, 0.5 metres in width, and 1.5 metres in height, while the subdomain has dimensions of 4.22 metres in length, 0.6 metres in width, and 0.17 metres in height.

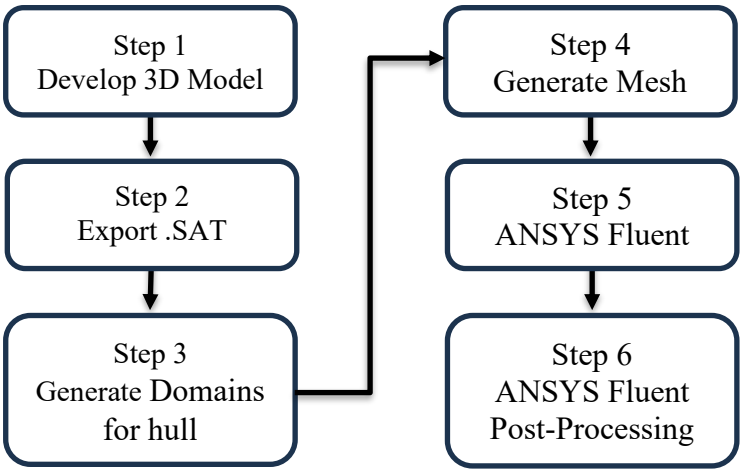


Figure 1. Steps of simulating the hull's resistance.

- Step 4: Generate mesh

The development of fine mesh is a crucial stage in the modeling of hull resistance because it allows the solution to generate more precise results. A suitable machine could produce the computational time, but it mainly relies on the computer's CPU. The primary domain's intake is positioned to face the bow, while the hull's outflow is positioned to face the stern. Body sizing on the global domain and face sizing on the hull are the two-mesh sizing employed. The meshing parameters are listed in Table 1.

Table 1. Meshing parameters of the simulation

Mesh Properties	CFD
Solver Preference	**Fluent**
Element Order	**Linear**
Element Size	**Default (225.92 Mm)**
Body Sizing	
Type	**Body Of Influence**
Element Size	**30.0 Mm**
Growth Rate	**Default (1.2)**
Face Sizing	
Type	**Element Size**
Element Size	**15.0 Mm**
Proximity Size Function Source	**Faces And Edges**
Statistics	
Nodes	**102612**
Elements	**540272**

- Step 5: ANSYS Fluent setup

In this step, the solution includes a pressure-based solver type, an absolute velocity formulation function, a constant time condition, and a gravitational acceleration. In addition, the viscous model is set to k-epsilon for the model set. Finally, the simulation

applied the near-wall treatment of typical wall functions and a realisable k-epsilon model. The realisable k-epsilon model has an advanced formulation for turbulence compared to the k-epsilon model, thus providing more accurate results.

The multiphase model is otherwise configured to fluid volume with open channel flow and an implicit body first formulation. Water and air are the two fluid materials used in the material settings, with a water density of 1000 kg/m3. They were chosen for this multiphase model because the hull is partially immersed in the body of water where the procedure is carried out. Fiberglass is chosen as the solid material for the research.

For boundary conditions, five velocities in the inlet towards the bow at 1.39 m/s (5 km/h) to 1.95 m/s (7 km/h) were evaluated against the hull to determine the force acting on it. Standard initialisation, automatic time-step approach, 0.5-time scale vector, and 250 iterations are used in the solution.

- Step 6: ANSYS Fluent Post-Processing

  In Fluent Post Processing, the iso-surface and vector are generated by the force acting on the hull and can be determined using the calculator function. This is because the resistance on the hull is similar to the resistance when the hull is moving forward.

2.2.2. CFD Analysis of BlueRobotics T200 Propeller

The self-propelling propeller is analysed using CFD analysis in this research to confirm the data supplied by the propeller performance chart. In addition, the thrust force generated at a given rotational speed may be calculated using the propeller's CFD analysis. Finally, the results of the CFD study will be compared to the thruster data given by BlueRobotics to support the thruster selection for the portable oil spill skimmer. The ANSYS Fluent T200 propeller is detailed here.

- Step 1: Conversion of the 3D model of BlueRobotics T200 to the (.STEP) format.
- Step 2: Export to the ANSYS Program in the SAT format.
- Step 3: Generation of domains for propeller

  The propeller is centred at the top of the page, and the domain is drawn around it. The enclosure and the revolving body are the two domains. The rotating body is a cylinder with an 85 mm diameter and a 32 mm thickness covering the propeller. For the examination of the spinning body, enclosure, and propeller, there are two subtracting Booleans.

Table 2. Meshing properties of propeller analysis

Mesh Properties	**CFD**
Solver Preference	**Fluent**
Element Order	**Linear**
Element Size	**Default (30.619 mm)**
Body Sizing	
Type	**Element Size**
Element Size	**20.0 Mm**
Capture Curvature	**No**
Statistics	
Nodes	**102612**
Elements	**540272**

- Step 4: Mesh Generation

https://doi.org/10.31436/iiumej.v26i3.3253

The propeller is subjected to face sizing with a 20 mm element size for this study. On the Z-axis, the propeller, intake, and outlet rotate in a clockwise direction. The default mesh size for the spinning body and enclosure is 30.619 mm. Table 2 shows the meshing characteristics.

- Step 5: ANSYS Fluent setup
  - The solution in this work uses a pressure-based solver type, an absolute velocity formulation function, a transient time condition, and gravitational acceleration. Additionally, the viscous model is set to the k-epsilon model in the model settings.
  - The water used in the CFD analysis has a 1000 kg/m3 density.
  - The rotating body zone is set to mesh in the cell zone condition setting. The motion end propeller is configured to rotate at 3000 RPM, and the intake is subjected to a velocity of 3.8 m/s for the boundary condition.
  - It is set to hybrid initialisation for the solution, which means the software guesses the variables. With 100 time steps and a maximum iteration of 10, the time step size is set to 0.05 seconds.
- Step 6: Post Processing

  Streamline is created as the container to represent the flow pattern in Fluent Post Processing. The propeller's force may be calculated using the calculator tool. When the propeller rotates in its blade direction, the created force is the benefit of distress.

2.3. Manoeuvrability Testing

The analysis of maneuverability was conducted using a series of tests at different distances (m). Table 3 shows the expected results of the analysis. A Knots unit was used as the POSS traveled on the water surface. The speed of the portable oil spill skimmer varies due to the external factors of wind and water currents.

Table 3. Expected result of maneuverability testing

Test	Distance (m)	Time Taken (s)	Speed (m/s)	Knots
1	**30**			
2	**60**			
3	**100**			
	Average speed			

2.4. CFD Analysis of Oil Sloshing

The study framework provides a comprehensive explanation of the methodology employed for liquid analysis. The liquid sloshing investigation will use SolidWorks Computational Fluid Dynamics (CFD) simulation. Table 4 displays the sequential procedure of the computational fluid dynamics (CFD) simulation. The following instructions must be followed to conduct this simulation.

https://doi.org/10.31436/iiumej.v26i3.3253

Table 4. Step-by-step process of the oil sloshing analysis

Pre-processor
1. Define Solid Model Geometry
2. Define Gravity Axis and Motion Axis
3. Define Gases and Fluid Types
4. Define the Initial Condition and Substance Concentration
5. Global Mesh
Solution
6. Define the Global Goals
7. Select Transient Explorer Parameters
8. Run the Simulation
Postprocessor
9. Plot, View, and Export the Results
10. Compare and Verify the Results

2.4.1. Define Solid Model Geometry

The basic steps were conducted in the pre-processor. The solid model geometry from the SolidWorks software was converted into the Ansys software. Then, the element type was defined according to the analysed element. Following the conversion of the solid model, the appropriate element type was determined based on the characteristics and requirements of the specific analysis being conducted. This step ensures that the model is accurately represented and analysed using the designated elements within the Ansys software.

2.4.2. Gravity Axis and Motion Axis

To obtain precise data, it is necessary to establish the gravity and motion axes with the correct value for the simulation. The sequential process for defining the gravity and motion axis is as follows:

- The oil tank's gravity must be specified on the y-axis. Gravity has a constant value of -9.81 m/s2 since it exerts a continuous force towards the Earth's center.

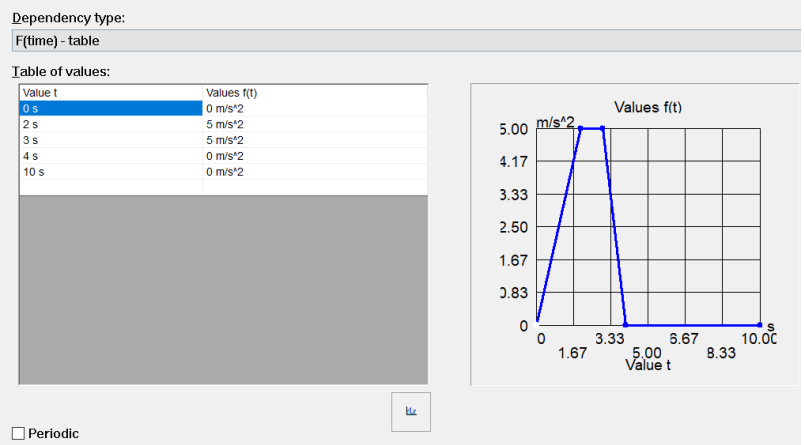


Figure. 2. Motion defined with a time variable

- The X-axis is the motion axis that indicates the direction of movement for the oil spill skimmer. The movement is defined by a variable that changes over time, as seen in Figure 2. The oil spill skimmer starts operating from a fixed position. The oil spill skimmer starts accelerating at 2 seconds, increasing at a constant rate of 5 m/s^2 until it reaches 3 seconds.

https://doi.org/10.31436/iiumej.v26i3.3253

The oil spill skimmer decelerates and stops completely after 4 seconds, with an acceleration of 0 m/s^2. The oil spill skimmer stops and stays still for 10 seconds.

2.4.3. Gases and Fluid Types

Air and propane are selected as the components for the simulation, namely in the states of gases and fluids. The oil tank consists of two main parts: the lower section is for liquid propane, and the upper section is for gases. Propane acts as a replacement for oil. Figure 3 illustrates the selected gases and fluids.

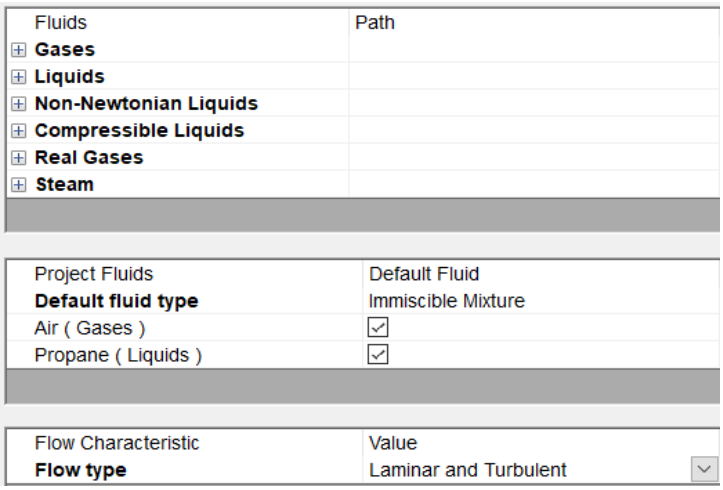
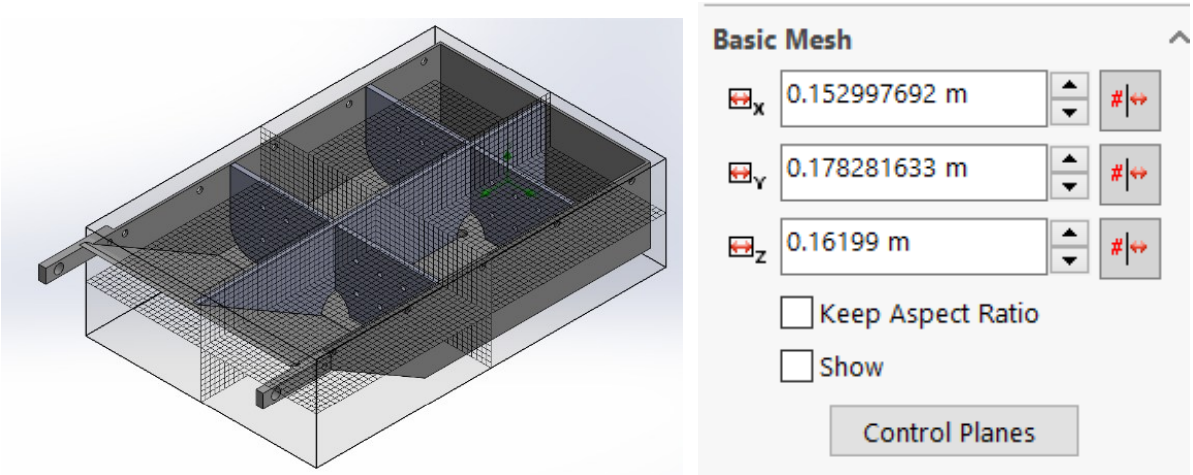


Figure 3. Selected gases and fluids

2.4.4. Meshing

Before simulating the design, meshing must be applied. The configuration for meshing in the flow simulation is displayed in Figure 4(a). The basic mesh refinement level is multiplied by two to obtain more accurate results, as shown in Figure 4 (b). Increasing the level of meshing detail directly correlates with an increase in the accuracy of the resulting data.



(a) (b)

Figure 4. (a) Meshing detail of oil tank, (b) the basic mesh values

https://doi.org/10.31436/iiumej.v26i3.3253

3. RESULT AND DISCUSSION

3.1. Result of CFD Analysis on Hull

The hull design of the Portable Oil Spill Skimmer was analysed using ANSYS Fluent. The findings were obtained after 250 iterations to achieve satisfactory solution convergence. Increasing the number of iterations leads to more precise computational fluid dynamics (CFD) findings. Five distinct examples have been examined, and the outcomes are presented in Table 5. According to the CFD analysis, the hull experiences a drag force of 2.7312 N when moving forward in the water at a speed of 1.39 m/s, or approximately 5.0 km/h. Figure 5 shows the increasing trend and the relationship between the drag force and the hull speed. Meanwhile, an increase of 2 km/h from 7.0 km/h produces a drag of 13.5542N on one hull, showing a rise of 396.27%, almost 5 times the drag force value at a speed of 5.0 km/h. These drag forces show increments when the speed of the Portable Oil Spill Skimmer increases.

Table 5. The relationship of drag force against the hull speed

Run	Velocity (m/s)	Velocity (km/h)	Drag Force of one hull (N)	Drag Force on two hulls (N)
1	1.39	5	2.7312	5.4624
2	1.53	5.5	3.5844	7.1688
3	1.67	6	9.2366	18.4732
4	1.81	6.5	12.1222	24.2444
5	1.95	7	13.5542	27.1084

The water flow pattern was shown in Figure 6 when moving forward at the speed of 1.95 m/s. The wake velocity generated at the stern of the hull increases with the hull speed. In addition, a minor wave hits the hull's bow at 1.39 m/s, becoming more aggressive when the speed increases to 1.81 m/s and 1.95 m/s. In this case, the skimmer utilizes a catamaran with two hulls side by side, resulting in a total drag force of 27.1084 N. Hence, the two thrusters must generate forward thrusts of more than 27.1084 N to achieve the hull speed of 1.95 m/s.

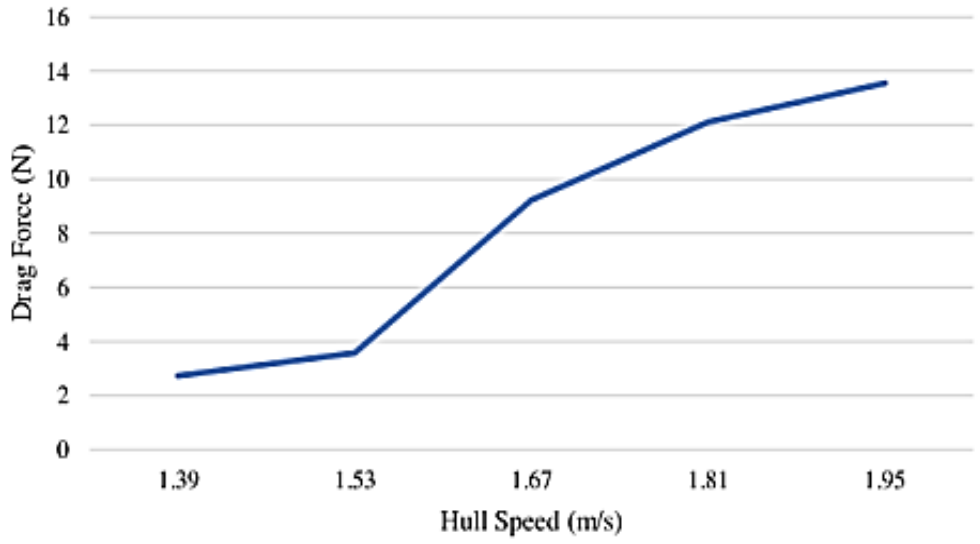


Figure 5. Relationship between the drag force and hull speed

https://doi.org/10.31436/iiumej.v26i3.3253

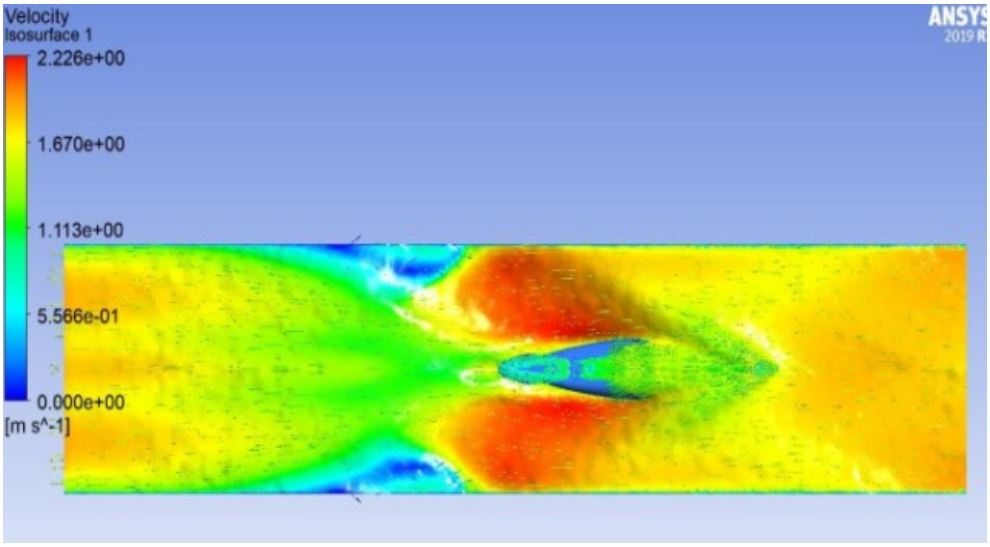


Figure 6. Flow pattern of water at a speed of 1.95 m/s

3.2. CFD Analysis on BlueRobotics T200 Propellers

From the CFD analysis that has been performed on the self-propelling propeller, it is found that the propeller can generate a thrust force of 36.8246 N at 3000 rpm and an inlet velocity of 3.8m/s. This thrust value is obtained from the ANSYS Fluent Post Processor of results, namely the function calculator, which is the function of thrust force, and the location is the propeller selection, as shown in Figure 7. In addition to that, the pressure contour in Figure 8 shows the maximum pressure of 5.109e^{+04} Pa and a minimum pressure of -1.284e^{+05} Pa on the propeller blades.

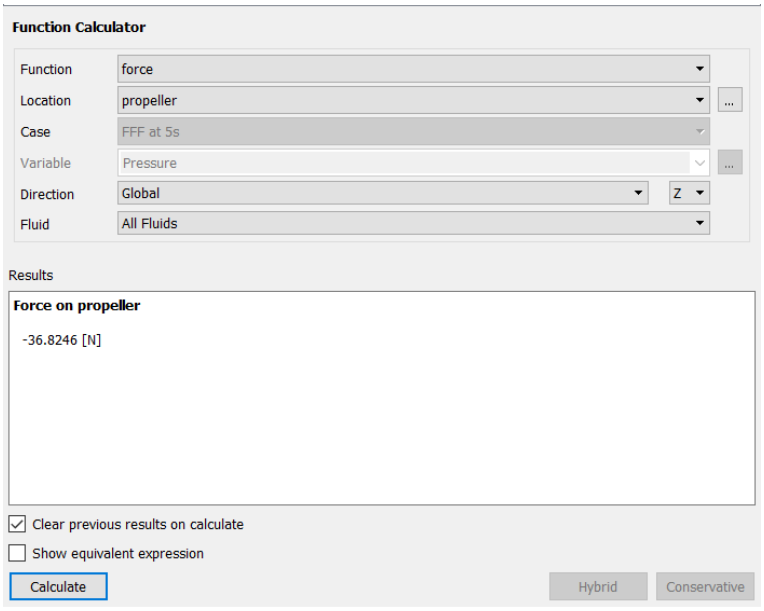


Figure 7. The value of thrust force exerted on the propeller.

https://doi.org/10.31436/iiumej.v26i3.3253

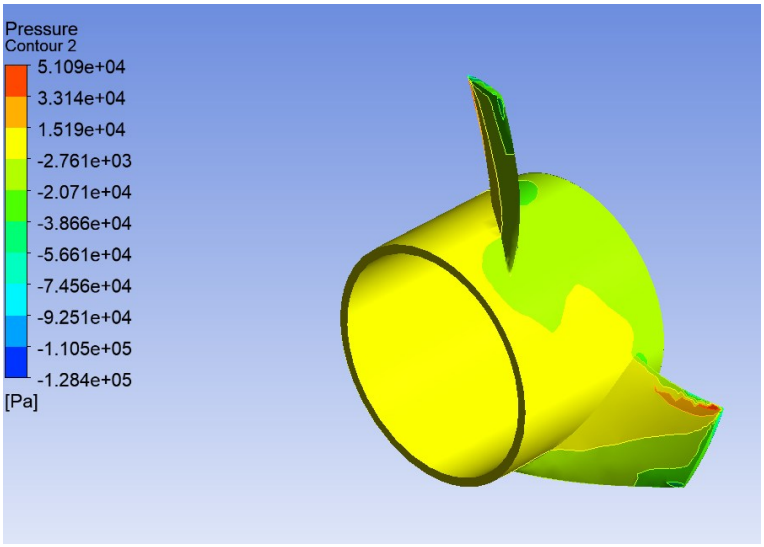


Figure 8. The pressure contour on the propeller blades

Furthermore, data provided by the propeller manufacturer, BlueRobotics, reveals that the T200 model can generate a maximum thrust of 3.71 kg or 36.383 N. This thrust level is achieved when the thruster rotates at a speed of 2995 rpm and is powered by a 12 V supply. Additionally, the results of the Computational Fluid Dynamics (CFD) simulation conducted on the propeller indicate a slight difference of 1.21% in the thrust value, comparing the supplier data (36.38 N) with the CFD result (36.4 N). This suggests that the setup parameters used in the CFD calculations are accurate and reliable. Based on the design of the Portable Oil Spill Skimmer, there are two T200 thrusters installed at the rear part of the frame to facilitate its movement on water. With these thrusters, the combined thrust force is 72.8 N, which is capable of overcoming the hull drag of 27.1084 N to propel the vehicle up to a speed of 1.95 m/s or even higher, provided the water conditions are ideal. These data prove that the T200 thrusters used on the Portable Oil Spill Skimmer are reliable.

3.3. Result of POSS Manoeuvrability

The portable oil spill skimmer consists of propellers capable of self-propulsion. Figureure 9 shows the portable oil spill skimmer manoeuvrability testing at Tasik Ayer Keroh. The testing was conducted by measuring the time the portable oil spill skimmer took to reach the desired distance. For instance, the distance for the portable oil spill skimmer is 100 meters; thus, the time taken to get the 100-meter distance is observed. A total of 3 tests were conducted with different distances of 30, 60, and 100 meters. The data collected for the manoeuvrability testing is shown in Table 6. Equation 1 shows the conversion from meters per second (m/s) to Knots. The result indicates that the POSS was able to be manoeuvred at a speed of 3.72 Knots, which adheres to the typical speed for an unmanned surface vehicle (USV) of 3 Knots and above [23]-[25]. The water currents and wind influenced the speed of the POSS during the experiment process.

$$\text{Knots} = \text{meters per second} \times 1.943844 \quad (1)$$

where the speed in knots is equal to the speed in meters per second multiplied by 1.943844, then we can obtain Knots = 1.88 × 1.943844 = 3.65.

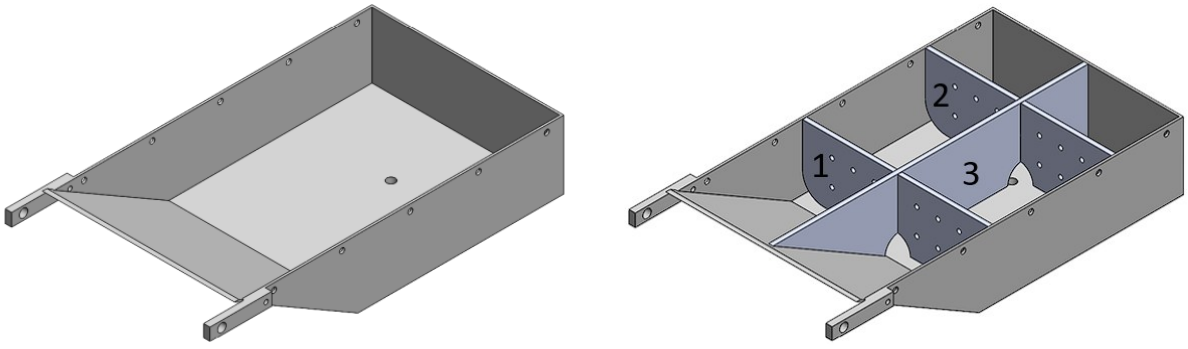
Figure 9. Portable oil spill skimmer manoeuvrability testing at Tasik Ayer Keroh

Table 6. The data collection of the POSS speed

Test	Distance (m)	Time Taken (s)	Speed (m/s)	Knots
1	30	16	1.88	3.65
2	60	31	1.94	3.77
3	100	52	1.92	3.73
Average speed			1.91	3.72

3.4. Result of Oil Sloshing

The simulation was conducted on the oil tank design to prove that oil sloshing could be reduced. The design shape of the oil tank without a baffle is shown in Figure 10(a), while the oil tank design with a baffle is shown in Figure 10(b). The design was developed with a size of 773 x 450 mm, and a total of 3 baffles was implemented. Based on Figure 10(b), baffles 1 and 2 were designed to be parallel to each other, with baffle 3 positioned in the middle. Baffle 3 was explicitly designed to reduce the oil sloshing during left or right turns, and baffles 1 and 2 were designed to reduce the oil sloshing during acceleration and deceleration.



(a) (b)

Figure 10. The design shape of the oil tank (a) without baffles, (b) with baffles

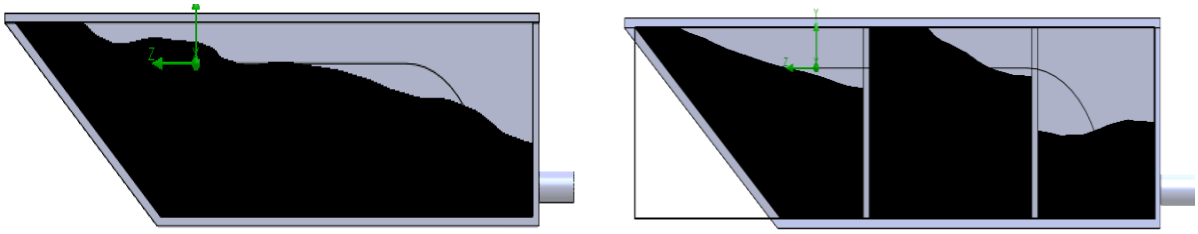
3.4.1. Liquid Sloshing Analysis on the Designed Oil Tank with and without baffles

The simulation was conducted using an oil tank design with and without baffles to investigate the difference in the impact of oil on the tank. The simulation measured and

https://doi.org/10.31436/iiumej.v26i3.3253

collected the torque and force acting upon the oil tank. Based on Figure 11 (a), it was observed that the time taken for the oil to stabilise in the oil tank without a baffle was more than 10s. The maximum torque exerted on the oil tank was 1195.43 Nm, as shown in Figure 11 (a). This indicates that the POSS has a risk of accidents and overturns due to the ample torque and force exertion.

Figure 12 (b) shows the liquid sloshing effect in the designed oil tank with baffles. The sloshing effect lasts only 5 seconds before the oil stabilises in the tank. This indicates that implementing a baffle in the oil tank has a reduced impact on the performance of the designed oil tank. The effect of the torque exerted on the oil tank is only 392 Nm, a reduction of 67.21%, minimizing the risk of the POSS overturning, as shown in Figure 12(b). This proves that implementing several baffles can eliminate the pitching and heaving effect and provide a safer working environment.



(a) (b)

Figure 11. Liquid sloshing effect in oil tank without baffle

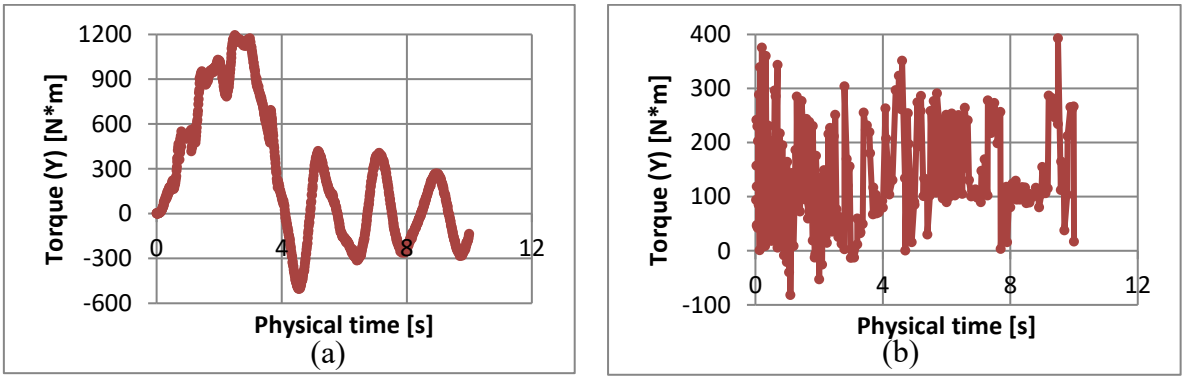


Figure 12. (a) The torque force of the sloshing effect without baffles, (b) The torque force of the sloshing effect with baffles

The result obtained from the oil sloshing analysis provides an essential understanding of the significance of utilising baffles when designing the oil tank of a motorised vehicle. Without the implementation of a baffle, the impact on an oil tank increases the risk of damage or failure to the tank [5], causing imbalance and potential damage to the POSS [10], and making it challenging to keep the POSS in control. To resolve this issue, baffle is implemented to reduce the impact caused by oil sloshing. The analysis results prove that designing an oil tank with baffles significantly reduces the effect of oil sloshing compared to an oil tank without baffles.

4. Conclusion

The Portable Oil Spill Skimmer (POSS) was designed to enhance the mechanical technology used in Oil Spill Response and Recovery (OSRR) operations. The developed POSS consists of an oil tank to recover oil and propellers for self-propelling capability. The POSS

could travel at a speed of 3.72 Knots, adhering to the standard of USV specification. However, during the POSS experimental testing, the POSS experiences instability caused by the sloshing effect. The sloshing effect occurs when recovered oil is present in the oil tank. The instability increases when the oil tank is partially filled. The impact of oil sloshing damages the structure of the oil tank and the POSS's stability. Thus, baffles were implemented to reduce the sloshing effect on the POSS. The oil sloshing was analysed by comparing the impact of sloshing, with and without the implementation of baffles. The finding shows significant improvement where the oil sloshing torque force was reduced by 67% from 1195.4N to 392N with the implementation of 3 baffles. Increasing the number of baffles reduces the sloshing impact but increases the weight of POSS. Only 3 baffles were installed in the oil tank to maintain the portability feature of POSS. This paper successfully proved that the sloshing effect was reduced by implementing baffles, enhancing the safety feature of the POSS.

ACKNOWLEDGEMENT

The authors would like to acknowledge the financial support from Universiti Teknikal Malaysia Melaka (UTeM) through the research grant with reference no. PJP/2020/FKP/PP/S01776 for all the support.

REFERENCES

[1] Abdullah L, Jamaludin Z, Chiew TH, Rafan NA, Yuhazri MY. (2012) Extensive Tracking Performance Analysis of Classical feedback control for XY Stage ballscrew drive system. Applied Mechanics and Materials, 30(229):750-5. https://doi.org/10.4028/www.scientific.net/AMM.229-231.750

[2] Tantichukiad K, Yahya A, Mustafah AM, Mohd Rafie AS, Mat Su AS. (2024) Computational Fluid Dynamics Evaluations on New Designs of the Delta-Shaped Blade Darrieus Hydrokinetic Turbine. International Journal of Automotive and Mechanical Engineering (IJAME), 21(2):11247–11262. https://doi.org/10.15282/ijame.21.2.2024.6.0869

[3] Ali MA, Idayu N, Salleh MS, Sivarao S, Yamaguchi M. (2022) Optimization of injection machine parameters on high tensile strength of polypropylene/fibreglass. Journal of Advanced Manufacturing Technology (JAMT), 20(16):3.

[4] Aiman TMKTK, Thamrin NM, Hanif M, Ramli M. (2023) Underwater brushless motor performance analysis of twin hull-based unmanned surface vehicle for water quality monitoring. IEEE 6th International Conference on Electrical, Electronics and System Engineering (ICEESE), 20–25. https://doi.org/10.1109/ICEESE56169.2023.10278188

[5] Zhang E, Chen H, Wang L, Huang J, Xu F. (2024) Numerical analysis of oil liquid sloshing in fuel tanks of plug-in hybrid electric vehicles under different structure baffles. Journal of Physics: Conference Series, 2879(1):012024. https://doi.org/10.1088/1742-6596/2879/1/012024

[6] Liu Z, Y Feng Y, Lei G, Li Y. (2019) Hydrodynamic performance in a sloshing liquid oxygen tank under different initial liquid filling levels. Aerospace Science and Technology, 85:544–555. doi: 10.1016/J.AST.2019.01.005.

[7] Zheng XL, Zhang H, Ren YY, Wei ZH, Song XG. (2017) Rollover stability analysis of tank vehicles based on the solution of liquid sloshing in partially filled tanks. Advances in Mechanical Engineering, 9(6). doi: 10.1177/1687814017703894.

[8] Toumi M, Bouazara M, Richard MJ. (2009) Impact of liquid sloshing on the behaviour of vehicles carrying liquid cargo. European Journal of Mechanics - A/Solids, 28(5):1026–1034. doi: 10.1016/J.EUROMECHSOL.2009.04.004.

[9] Zhang E, He R. (2020) Numerical analysis on the sloshing of free oil liquid surface based on fuel tanks of different shapes. Journal of Mechanical Engineering Science, 234(18):3584–3599. doi: 10.1177/0954406220916483.

[10] Wu X, He R. (2024) Numerical study on characteristics and mechanism of fuel sloshing in automobile fuel tank. Journal of the Brazilian Society of Mechanical Sciences and Engineering, 46(1):1–16. doi: 10.1007/S40430-023-04601-3/METRICS.

[11] Kim J, Kim Y, Kim SY, Lee KM, Sung YJ. (2020) Experimental Study of Slosh-induced Loads on an LNG Fuel Tank of a Container Ship. International Journal of Offshore and Polar Engineering, 30(02):141–148. doi: 10.17736/IJOPE.2020.JC766.

[12] Gurusamy S. (2023) Damping of liquid sloshing by floating balls. Physics of Fluids, 35(11). doi: 10.1063/5.0165870/2922927.

[13] Luo M, Su X, Lin P, Khayyer A, Zhao X. (2022) Investigation of Two-layer Liquid Sloshing by Using the Consistent Particle Method. International Journal of Offshore and Polar Engineering, 32(01):7–15. doi: 10.17736/IJOPE.2022.JC848.

[14] Liu Z, Feng Y, Liu Y, Lei G, Li Y. (2020) Fluid sloshing dynamic performance in a fuel storage tank under sinusoidal excitations. Applied Thermal Engineering, 168(4):114814. doi: 10.1016/J.APPLTHERMALENG.2019.114814.

[15] Liu D, Wu Y, Lin P. (2022) An experimental study of two-layer liquid sloshing under pitch excitations. Physics of Fluids, 34(5). doi: 10.1063/5.0093716/2846755.

[16] Yu L, Xue MA, Jiang Z. (2020) Experimental investigation of parametric sloshing in a tank with vertical baffles. Ocean Engineering, 213:107783. doi: 10.1016/J.OCEANENG.2020.107783.

[17] Nasar T, Sannasiraj SA. (2019) Sloshing dynamics and performance of porous baffle arrangements in a barge carrying liquid tank. Ocean Engineering, 183:24–39. doi: 10.1016/J.OCEANENG.2019.04.022.

[18] Tang YY, Liu YD, Chen C, Chen Z, He YP Zheng MM. (2021) Numerical study of liquid sloshing in 3D LNG tanks with unequal baffle height allocation schemes. Ocean Engineering, 234:109181. doi: 10.1016/J.OCEANENG.2021.109181.

[19] Zang Q, Liu J, Yu L, Lin G. (2019) Boundary element analysis of liquid sloshing characteristics in axisymmetric tanks with various porous baffles. Applied Ocean Research, 93:101963. doi: 10.1016/J.APOR.2019.101963.

[20] Ünal UO, Bilici G, Akyıldız H. (2019) Liquid sloshing in a two-dimensional rectangular tank: A numerical investigation with a T-shaped baffle. Ocean Engineering, 187:106183. doi: 10.1016/J.OCEANENG.2019.106183.

[21] Huang P. (2023) A comprehensive investigation on liquid sloshing of rectangular water tank with vertical baffles. Ocean Engineering, 288:116126. doi: 10.1016/J.OCEANENG.2023.116126.

[22] Jamalabadi MYA, Ho-Huu V Nguyen TK. (2018) Optimal design of circular baffles on sloshing in a rectangular tank horizontally coupled by structure. Water 10(11). doi: 10.3390/w10111504.

[23] Nagalakshmi PSS, Vijaya N, Ibrahim SM. (2023) Entropy Generation of Three-Dimensional Williamson Nanofluid Flow Explored with Hybrid Carbon Nanotubes over a Stretching Sheet. CFD Letters, 15 (7):112-130. https://doi.org/10.37934/cfdl.15.7.112130

[24] Chi C, Pallayil V, Chitre M. (2020) Design of an adaptive noise canceller for improving performance of an autonomous underwater vehicle-towed linear array. Ocean Engineering, 202:106886. doi: 10.1016/J.OCEANENG.2019.106886.

[25] Zhang H, Zhang S, Wang Y, Liu Y, Yang Y, Zhou T, Bian H. (2020) Subsea pipeline leak inspection by autonomous underwater vehicle. Applied Ocean Research, 1(107):102321. doi: 10.1016/J.APOR.2020.102321.

https://doi.org/10.31436/iiumej.v26i3.3603

FORMATION OF VOID IN BFS/$CACO_3$ DIFFUSION COUPLE

AHMAD ABDUL MUN'IM ISMAIL[1*], ALYA NAILI ROZHAN[1], MUHAMMAD RAFIQ HAIKAL ROSDIN[1], HADI PURWANTO[2], ABD MALEK ABDUL HAMID[1], M. H. ANI[1]

[1] Department of Manufacturing and Materials, Kulliyyah of Engineering, International Islamic University Malaysia, Jalan Gombak, 53100 Kuala Lumpur, Malaysia.
[2] Department of Engineering Management, Universitas Internasional Semen Indonesia, Kompleks PT. Semen Indonesia (Persero) Tbk. Jl. Veteran, Gresik Jawa Timur, 61122 Indonesia.

**Corresponding author: imismalz@gmail.com*

(Received: 31 January 2025; Accepted: 28 May 2025; Published online: 9 September 2025)

ABSTRACT: This study investigates void formation in the blast furnace slag (BFS) and calcium carbonate ($CaCO_3$) diffusion couple, which is critical for understanding the interdiffusion process in cement production. The experimental analysis involved high-temperature diffusion experiments, focusing on the volume fraction of void at the BFS/$CaCO_3$ interface, the activation energy of void formation, and the I-V measurement of void formation at the BFS/$CaCO_3$ interface. Void measurements revealed a 25% increase after exposure to the specified temperature, while the activation energy for void formation was calculated to be – 41.48 kJ/mol. I-V measurements revealed ionic diffusion as the dominant mechanism for void formation, with an average decomposition rate of 1.4598×10^{-12} m^2s^{-1}. These findings provide valuable insights for utilizing BFS in cement production.

ABSTRAK: Kajian ini menyelidik pembentukan rongga dalam pembentukan sanga relau letupan (BFS) dan kalsium karbonat ($CaCO_3$), di mana ianya sangat penting bagi memahami proses antara resapan (interdiffusion) dalam pengeluaran simen. Analisis eksperimen melibatkan ujian resapan pada suhu tinggi, dengan tumpuan kepada pecahan isipadu rongga antara permukaan BFS/$CaCO_3$, tenaga pengaktifan bagi pembentukan rongga, dan pengukuran I-V bagi pembentukan rongga antara muka BFS/$CaCO_3$. Pengukuran rongga menunjukkan peningkatan sebanyak 25% selepas terdedah kepada suhu yang ditetapkan, manakala tenaga pengaktifan bagi pembentukan rongga dikira sebanyak –41.48 kJ/mol. Pengukuran I-V menunjukkan bahawa penyebaran ionik merupakan mekanisme dominan bagi pembentukan rongga, dengan kadar penguraian purata sebanyak 1.4598×10^{-12} m^2/s. Dapatan ini memberikan pemahaman penting bagi penggunaan BFS dalam pengeluaran simen.

KEYWORDS: *void formation, diffusion couple, blast furnace slag, calcium carbonate*

1. INTRODUCTION

The study of void formation in the BFS/$CaCO_3$ diffusion couple explores the mechanisms underlying solid-state diffusion and the implications of this process on material properties. Blast Furnace Slag (BFS), a byproduct of the iron and steel industry, contains valuable chemical components, including calcium oxide (CaO) and silicon dioxide (SiO_2). Its utilization in cement production, particularly when combined with calcium carbonate ($CaCO_3$), presents an opportunity to reduce waste and improve resource efficiency[1], [2], [3]. However, understanding the diffusion behavior [4], [5] between these materials, particularly the

https://doi.org/10.31436/iiumej.v26i3.3603

formation of voids, is crucial to optimizing the production process and ensuring the structural integrity of the final product.

Solid-state diffusion, which occurs in materials with lattice imperfections, plays a pivotal role in the chemical reactions between BFS and $CaCO_3$ [6], [7]. The diffusion of calcium ions (Ca^{2+}) from $CaCO_3$ into BFS leads to the formation of calcium silicate phases such as alite (Ca_3SiO_5) and belite (Ca_2SiO_4)[8], [9]. These phases are critical in cement production as they contribute to the strength and durability of the material. However, the formation of voids during this diffusion process can compromise the material's mechanical properties, necessitating a thorough investigation into the factors influencing void formation[10], [11].

Previous studies have shown that the diffusion of ions such as Ca^{2+} is influenced by temperature, concentration gradients, and the presence of impurities [12], [13], [14]. The introduction of foreign elements, such as aluminum oxide (Al_2O_3) or iron oxide (Fe_2O_3), can further complicate the diffusion process by altering the reaction kinetics and contributing to void formation[15], [16]To better understand these phenomena, this research examines the diffusion behavior of BFS and CaCO3 under various temperature conditions, focusing on the relationship between ion migration and void formation.

Void formation during solid-state diffusion is a complex process influenced by the interaction of multiple variables. During these reactions, voids are frequently observed at the interface due to the solid-state diffusion of Ca^{2+} ions from $CaCO_3$ into BFS. Therefore, this study focused on the volume fraction of voids, their activation energy, and the role of alite and belite phases for understanding the BFS-$CaCO_3$ cement and the efficiency of the clinker formation process.

2. METHODOLOGY

2.1. Furnace Design and Setup

A high-temperature vertical furnace was employed for the diffusion experiments. The furnace's temperature was monitored using a Type R thermocouple, capable of measuring temperatures up to 1723 K, which was positioned at the center of the furnace to ensure accurate readings. An alumina tube, measuring 1000 mm long with an outer diameter of 60 mm and an inner diameter of 50 mm, served as the reaction chamber. The BFS and $CaCO_3$ samples were horizontally clamped together using wire gauze and placed within the furnace's isothermal zone, which was determined to be 100 mm long.

2.2. Sample Preparation

BFS and $CaCO_3$ were ground into fine powders using an alumina mortar and pestle to achieve homogeneity. The powders were then pressed into cylindrical pellets, each 10 mm in diameter and approximately 3 mm thick. These pellets were sintered in an air environment at 873 K for six hours. One side of each pellet was mechanically polished using P2000 grit emery paper to ensure optimal contact during the diffusion experiments. The polished BFS and $CaCO_3$ samples were then paired and clamped together with wire gauze to form the diffusion couple, as shown in Fig. 1. BFS and $CaCO_3$ samples were prepared and subjected to high-temperature conditions (1273 K, 1323 K, and 1373 K) for varying time periods (86.4 ks, 259 ks, and 432 ks).

https://doi.org/10.31436/iiumej.v26i3.3603

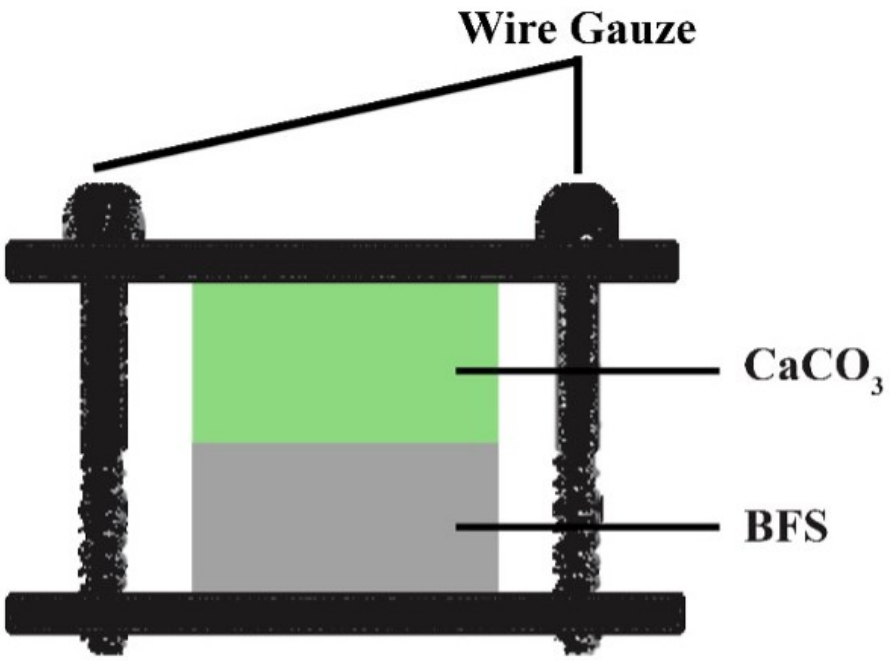


Figure 1. Schematic illustration of BFS/CaCO3 diffusion couple.

2.3. Characterization and Evaluation

The elemental composition of the BFS/$CaCO_3$ diffusion couple was analyzed before and after the diffusion process using X-ray Diffraction (XRD) and Scanning Electron Microscopy with Energy Dispersive X-ray analysis (SEM-EDX). XRD analysis, performed with a Bruker D2 Phaser using Cu Kα radiation, confirmed the composition of the samples before diffusion. Post-diffusion, SEM-EDX analysis (JEOL JSM-IT100) was conducted at 10 kV to investigate the elemental distribution across the sample's cross-section. The samples were prepared for SEM-EDX by mounting them in resin and sectioning with a diamond cutter to expose the cross-section.

2.4. Void Formation Calculation and Analysis

The diffusion coefficient of calcium ions (Ca^{2+}) from $CaCO_3$ into BFS was calculated using the Boltzmann-Matano method, which involves analyzing the concentration profile of the diffusing species. The concentration profiles were obtained from the SEM-EDX analysis and used to determine the diffusivity of Ca^{2+} ions. Additionally, the volume fraction of voids formed during the diffusion process was calculated using the formula proposed by [17] and [18] Where v_{void}^{ox} represents the formation of voids within the scale during the oxidation process, l pertains to the inner scale, and L corresponds to the outer scale with L > 1. $(L - l)^3$ is used to convert the void formation into cubic dimensions, mirroring the actual structure of the oxide scale diffused into the sample.

$$f_v = \frac{v_{void}^{ox}}{(L-l)^3} \quad (1)$$

2.5. Diffusion Couple I-V Experiment

The diffusion experiments involved setting up a tracer diffusion couple, where platinum wire electrodes were used for both working and counter electrodes due to their high melting point and chemical resistance. The experiment was conducted at 1373 K, and the potential difference between the electrodes was measured over time. Tafel plots were generated to

https://doi.org/10.31436/iiumej.v26i3.3603

evaluate the electrochemical reaction rate, which provided insights into the diffusion of ions, particularly Ca2+, within the BFS/CaCO3 system. The reaction rate was calculated using the Tafel slope, which quantitatively measures ion transport in the diffusion couple. [19], [20], [21]

This methodology outlines a comprehensive approach to studying the formation of voids in the BFS/CaCO3 diffusion couple. It uses SEM, EDX, XRD, and I-V experiments to provide detailed insights into the formation of voids impacted by diffusion behaviors.

3. RESULTS AND DISCUSSION

3.1. Volume Fraction of Voids In the BFS/$CaCO_3$ Interface

SEM images showed significant void formation at the BFS/$CaCO_3$ interface after prolonged exposure at high temperatures. Voids were observed as annealing time and temperature increased. The formation of voids in limestone, as depicted in Fig. 2, occurs due to the release of CO_2 gas into the atmosphere when exposed to high temperatures. In these conditions, CO_2 becomes volatile and escapes from the limestone. Several studies have explored the mechanism behind void formation in limestone. For instance, when limestone is heated, a chemical reaction occurs, releasing CO_2 gas as a byproduct [22]. The trapped gas creates internal pressure, and if this pressure cannot escape, voids or bubbles form within the material [23]. Additionally, higher temperatures accelerate the decomposition reaction, causing a more rapid release of CO_2 gas, which leads to pressure gradients that contribute to void formation [24]. In the case of BFS-limestone, void formation is driven by the Kirkendall effect, where uneven diffusion rates between species cause voids at the interface [25].

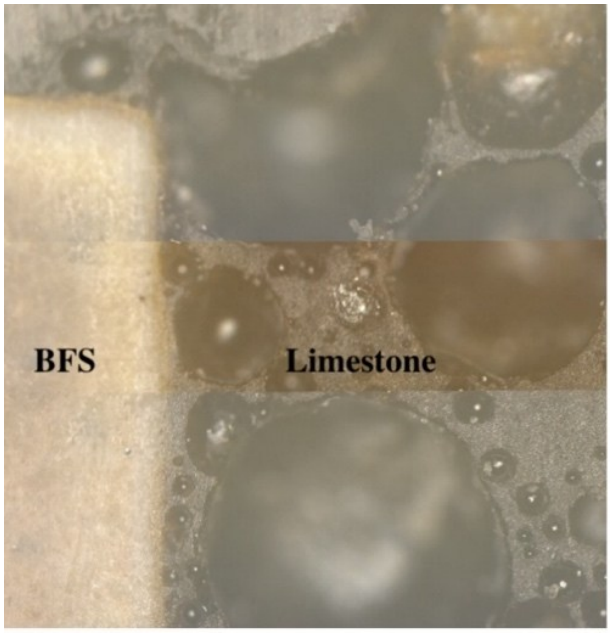


Figure 2. Formation of void in Limestone from one of the cross sections exposed to 1373 K at 432ks.

Fig. 3 shows the distribution of voids across the limestone, interlayer oxide, and BFS regions. The SEM image highlights a higher concentration of voids in the limestone area, with fewer voids observed in the interlayer oxide and BFS regions. To confirm this observation, the void percentage in each region can be calculated using the formula provided in Equation 1.

https://doi.org/10.31436/iiumej.v26i3.3603

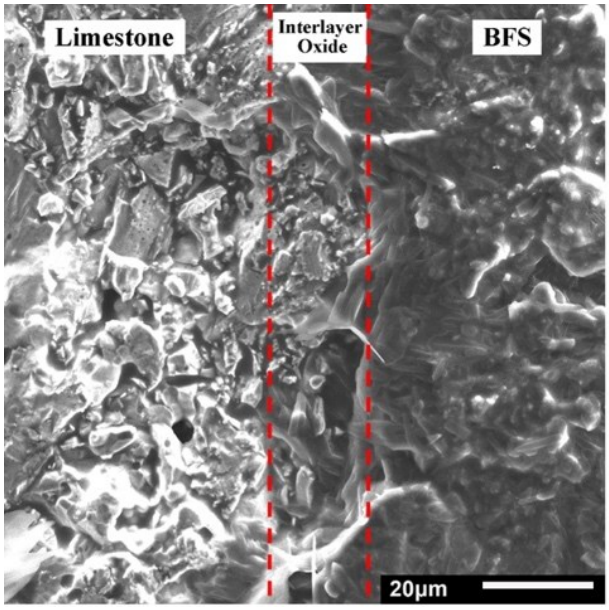


Figure 3. Cross-sectional micrograph of BFS/CaCO3 diffusion couple from one of the cross sections exposed to 1373 K at 432ks. The red dashed line indicates the interlayer oxide boundary.

Fig. 4 demonstrates that the volume fraction of voids decreases with longer exposure times, plateauing by 432 ks. This suggests that the $CaCO_3$ has fully reacted, CO_2 has been completely released, and no further voids are formed. In the early stages, the $CaCO_3$ actively releases CO_2, resulting in a higher void concentration at the BFS/$CaCO_3$ interface. Samples exposed to lower temperatures, such as 1273 K, exhibit fewer voids than those exposed to 1323 K and 1373 K. At the initial exposure time, the difference in void volume fraction reaches up to 30%. However, the void fractions for samples at 1323 K and 1373 K are pretty similar, as these temperatures approach the vaporization point of $CaCO_3$. The data show that the volume fraction of voids aligns well with the governing equation 1.

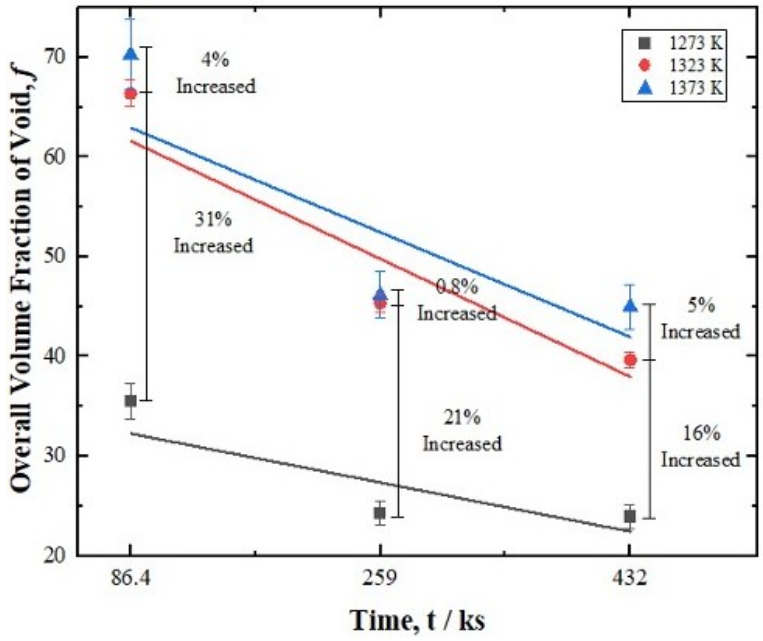
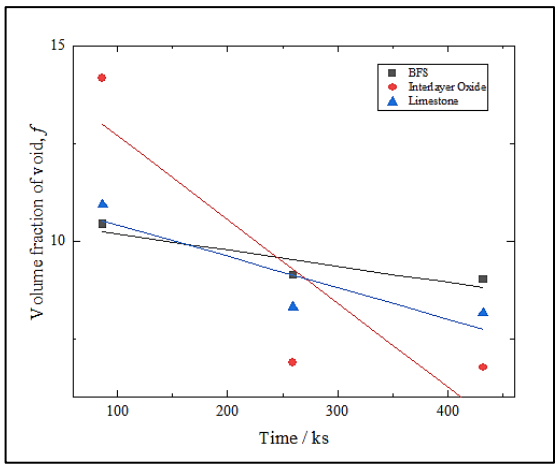


Figure 4. Overall Volume Fraction of Void vs Time

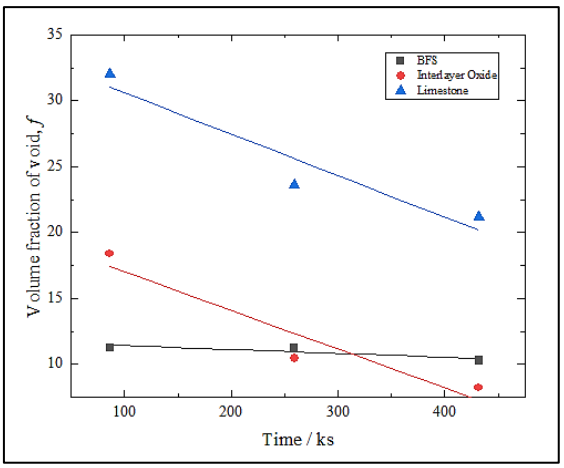
https://doi.org/10.31436/iiumej.v26i3.3603

Figure 4 also illustrates that higher temperatures increase void volume, ranging from 0.8% to 31%. As exposure time increases, void formation decreases and stabilizes. Overall, the void formation drops by 14%. The figure also suggests that increasing temperature will cause the void fraction to reach a steady state, implying that the void percentage may remain constant with prolonged exposure at higher temperatures.

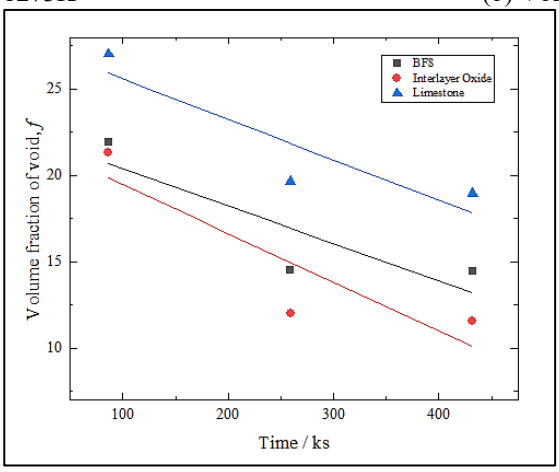
Further analysis of the void percentage in the BFS, interlayer oxide, and limestone regions, as shown in Figures 5 (a), (b), and (c), reveals a clear distinction in void formation. Figure 5 (a) shows that the reaction is still unstable; therefore, the voids at the three-phase boundary do not exhibit a clear pattern. As the temperature increases, the limestone region shows the highest volume of voids, followed by the BFS and the interlayer oxide, as seen in Figures 5 (b) and 5 (c). This is attributed to the decomposition of limestone, where the release of CO_2 at elevated temperatures generates more voids in the limestone region.



(a) Void at 1273K



(b) Void at 1323K



(c) Void at 1373K

Figure 5. Volume Fraction of Void at each layer vs Time at (a) 1273 K (b) 1323 K and (c) 1373 K

3.2 Activation Energy of Void Formation

Figure 6 illustrates the temperature dependence of the overall volume fraction of voids, showing a proportional increase that aligns with the trend in Figure 5. The void fraction is directly related to the rate of CO_2 volatilization and can be described using the Arrhenius equation. The void fraction, f, can be expressed using the Arrhenius equation as shown in Equation (2), where f_o is the pre-exponential function and Q is the void activation energy. The

https://doi.org/10.31436/iiumej.v26i3.3603

figure shows a linear relationship between log f and 1/T, representing the first reaction order. The activation energy for void formation, Q, was determined as – kJ/mol.

$$log\, f = log\, f_o - \frac{Q}{RT} \quad (2)$$

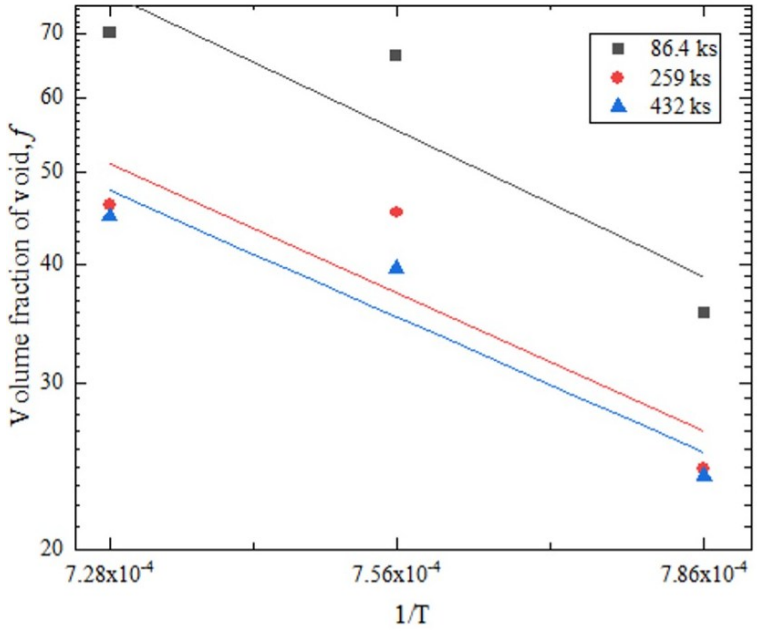


Figure 6. Overall Volume Fraction of Void vs Temperature

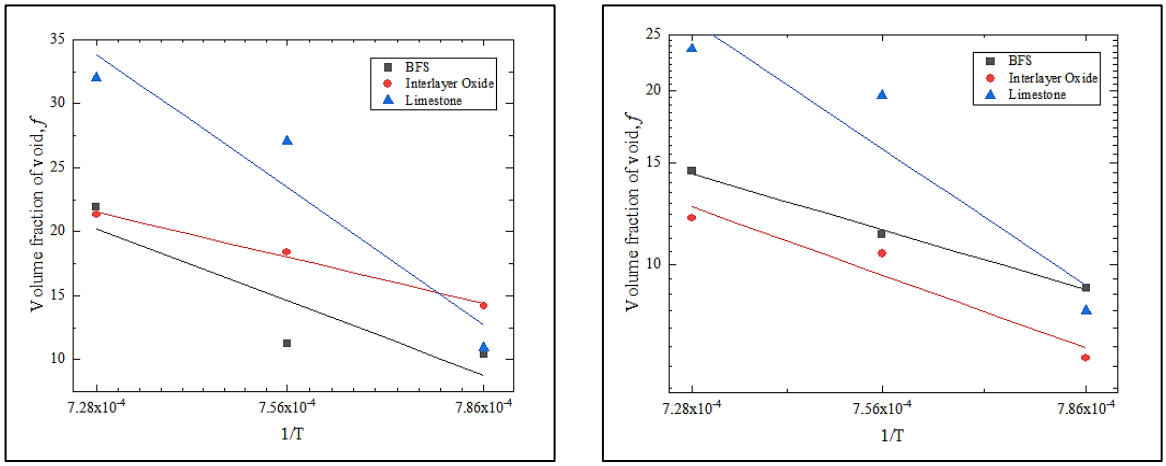
By plotting the logarithm of void volume fraction against inverse temperature, the activation energy for void formation was calculated. Table 1 shows the activation energy of void formation at the respective conditions. The longer the BFS/CaCO3 is exposed to the respective temperature, the energy for void formation at the respective couples is reduced. It shows the activation energy reduced by 7% at 432 ks.

Table 1. The activation energy for void formation, Q.

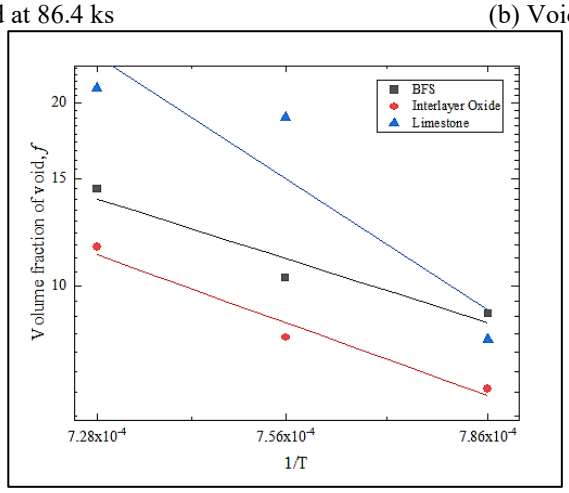
Reaction Time/ks	Activation Energy / kJ·mol^{-1}
86.4	−43.43
259	−40.90
432	−40.10

Further analysis can be done on the void fraction of BFS, interlayer oxide, and limestone, which can be referred to in Figures 7 (a), (b), and (c). A clear distinction can be observed based on this figure, where the highest volume of voids formed is in limestone, followed by BFS and interlayer oxide accordingly. This phenomenon confirms the reaction equation $CaCO_3(s) \rightarrow CaO\ (s) + CO_2\ (g)$, where CO_2 is gasified at high temperature. Thus, the volume fraction of void in the limestone area is the highest.

https://doi.org/10.31436/iiumej.v26i3.3603



(a) Void at 86.4 ks (b) Void at 259 ks



(c) Void at 432 ks

Figure 7. Volume Fraction of Void at each layer vs 1/T at (a) 86.4 ks (b) 259 ks and (c) 432 ks

3.3. I-V Measurement of Void Formation at BFS/$CaCO_3$ Interface

The I-V measurements were performed using platinum electrodes. The Tafel plot analysis revealed a linear relationship between the applied voltage and the current density, indicating that ionic diffusion is the dominant mechanism in void formation. The Tafel slope was used to calculate the reaction rate, which increased with temperature, confirming the temperature dependence of the diffusion process. Figure 8 presents the I-V plot of the BFS/$CaCO_3$ diffusion couple exposed to 1373 K for 60 s, 180 s, 300 s, 420 s, and 540 s. The I-V curve can be used to determine $E_{reaction}$ and $I_{reaction}$, which correspond to the potential and current from the main reaction in the diffusion couple. All results follow the Tafel pattern, allowing for the calculation of $E_{reaction}$ by finding the intersection points where a tangent line meets both the anodic and cathodic curves.

A notable observation is the rightward shift of the Tafel plot at 420 s compared to 300 s. This shift is due to the ongoing reaction between BFS and limestone, forming a diffusion oxide layer. As the reaction progresses, more calcium and oxygen are consumed, increasing the diffusion layer's formation. The rightward shift indicates that equilibrium has not yet been reached. Once the reaction involving the oxide layer is complete, the system undergoes another transition, as calcium and oxygen from the BFS/CaCO3 components continue to react. This leads to the leftward shift observed in the 540s Tafel plot, signifying that the calcium-oxygen reaction has become dominant and the system has reached a more stable state.

https://doi.org/10.31436/iiumej.v26i3.3603

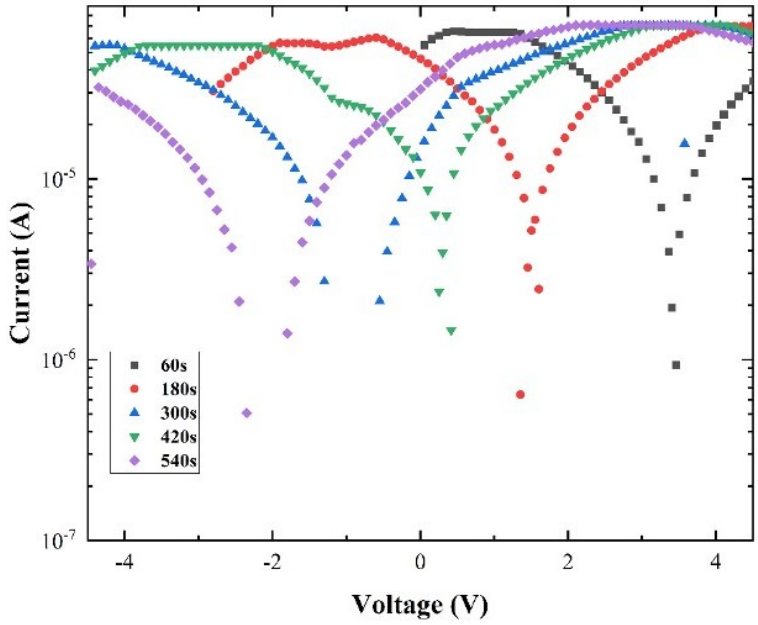


Figure 8. Tafel Plot of BFS/CaCO3 diffusion couple exposed to 1373 K at various time frames.

The $E_{reaction}$ from Figure 8 reflects the value from the decomposition of CaCO3. Over time, the decomposition rate is expected to decrease due to the reduced surface area, causing the curve to shift to the left. This leftward shift in the Tafel plot indicates a change in the dominant reaction within the diffusion couple, showing the system is reaching a more stable state. Additionally, the I-V plot can be used to calculate the Tafel slope and reaction rate, which are summarized in the decomposition rates in Table 2.

Table 2. Corresponding values calculated from the Tafel plot.

Time / s	$E_{reaction}$ / V	$I_{reaction}$ / A	Reaction rate (10^{-12}) / m^2s^{-1}
60	3.58	1.56 x 10^{-5}	1.6735
180	1.5	1.49 x 10^{-5}	1.5984
300	- 0.977	1.64 x 10^{-5}	1.7593
420	0.141	1.49 x 10^{-5}	1.5984
540	- 2.125	6.24 x 10^{-6}	0.6694

4. CONCLUSION

This study highlights the significance of void formation at the BFS/CaCO3 interface during high-temperature diffusion processes, which plays a crucial role in the interdiffusion process for cement production. The findings indicate a 25% increase in void volume after exposure to specific temperatures, with the activation energy for void formation calculated at –41.48 kJ/mol. Ionic diffusion was identified as the dominant mechanism for void formation, with an average decomposition rate of 1.4598 × 10^{-12} m^2s^{-1}. These insights are valuable for optimizing the use of BFS in cement production.

ACKNOWLEDGEMENT

This research was financially supported by the Ministry of Higher Education, Malaysia, with Grant Number FRGS/1/2019/TK07/UIAM/03/3.

REFERENCES

[1] T. Gao, L. Shen, M. Shen, L. Liu, and F. Chen, "Analysis of material flow and consumption in cement production process," *J Clean Prod*, vol. 112, pp. 553–565, Jan. 2016, doi: 10.1016/j.jclepro.2015.08.054.

[2] A. A. M. Ismail, M. R. H. Rosdin, A. M. A. Hamid, H. Purwanto, A. N. Rozhan, and M. H. Ani, "Producing Blast Furnace Slag Cement Clinker by Utilizing Redox Reaction Approach," 2023, pp. 89–96. doi: 10.1007/978-981-19-1851-3_8.

[3] A. A. M. Ismail *et al.*, "Blast Furnace Slag Cement Clinker Production Using Limestone-Hot Blast Furnace Slag Mixture," 2023, pp. 539–545. doi: 10.1007/978-981-19-9509-5_71.

[4] T. Tian, Y. Zhang, H. Zhang, K. Zhang, J. Li, and H. Wang, "Dissolution behavior of SiO2 in the molten blast furnace slags," *Int J Appl Ceram Technol*, vol. 16, no. 3, pp. 1078–1087, May 2019, doi: 10.1111/ijac.13120.

[5] M. Matsushima, S. Yadoomaru, K. Mori, and Y. Kawai, "Fundamental Study on the Dissolution Rate of CaO into Liquid Slag," *Tetsu-to-Hagane*, vol. 62, no. 2, pp. 182–190, 1976, doi: 10.2355/tetsutohagane1955.62.2_182.

[6] J. Zheng, X. Hu, Z. Ren, X. Xue, and K. Chou, "Solid-state reaction studies in Al2O3-TiO2 system by diffusion couple method," *ISIJ International*, vol. 57, no. 10, pp. 1762–1766, 2017, doi: 10.2355/isijinternational.ISIJINT-2017-042.

[7] S. Yasipourtehrani, S. Tian, V. Strezov, T. Kan, and T. Evans, "Development of robust CaO-based sorbents from blast furnace slag for calcium looping CO2 capture," *Chemical Engineering Journal*, vol. 387, p. 124140, May 2020, doi: 10.1016/j.cej.2020.124140.

[8] Y. A. Mohamed, A. E. M. O. Kasif, E. A. A. Alla, and M. M. Elmahadi, "Calculation of the formation process of clinker inside the rotary cement kiln," *Proceedings of the Voronezh State University of Engineering Technologies*, vol. 80, no. 1, pp. 233–239, Apr. 2018, doi: 10.20914/2310-1202-2018-1-233-239.

[9] S. Telschow, "Clinker Burning Kinetics and Mechanism," 2012.

[10] M. Li, S. Chen, H. Dai, H. Zhao, and B. Jiang, "Experimental Investigation on the Mass Diffusion Behaviors of Calcium Oxide and Carbon in the Solid-State Synthesis of Calcium Carbide by Microwave Heating," *Molecules*, vol. 26, no. 9, p. 2568, Apr. 2021, doi: 10.3390/molecules26092568.

[11] X. Zhang, W. J. Meng, and A. C. Meng, "Chemical potential gradient induced formation of Kirkendall voids at the epitaxial TiN/MgO interface," *Nanoscale*, vol. 15, no. 31, pp. 13086–13093, 2023, doi: 10.1039/D3NR01860A.

[12] A. Shankar, M. Görnerup, A. K. Lahiri, and S. Seetharaman, "Experimental Investigation of the Viscosities in CaO-SiO2-MgO-Al2O3 and CaO-SiO2-MgO-Al2O3-TiO2 Slags," *Metallurgical and Materials Transactions B*, vol. 38, no. 6, pp. 911–915, Dec. 2007, doi: 10.1007/s11663-007-9087-5.

[13] M. Chen, W. Zhang, Z. Zhao, T. Evans, and B. Zhao, "Effects of Minor Elements to the Liquidus Temperatures of Blast Furnace Slags," *ISIJ International*, vol. 56, no. 12, pp. 2156–2160, 2016, doi: 10.2355/isijinternational.ISIJINT-2016-268.

[14] Z. S. Ren, X. H. Hu, and K. C. Chou, "Model for diffusion coefficient estimation of calcium ions in CaO–Al $_2$ O $_3$ –SiO $_2$ slags," *Ironmaking & Steelmaking*, vol. 40, no. 8, pp. 625–629, Nov. 2013, doi: 10.1179/1743281212Y.0000000090.

[15] Y. Wang, X. Liu, H. Wang, and K. Vecchio, "The effect of oxides on Fe/Al interfacial reaction in Metal-Intermetallic Laminate (MIL) composites," *J Alloys Compd*, vol. 845, p. 156268, Dec. 2020, doi: 10.1016/j.jallcom.2020.156268.

[16] O. A. Bulavchenko *et al.*, "The Influence of Cu and Al Additives on Reduction of Iron(III) Oxide: *In Situ* XRD and XANES Study," *Inorg Chem*, vol. 58, no. 8, pp. 4842–4850, Apr. 2019, doi: 10.1021/acs.inorgchem.8b03403.

[17] M. Ueda, K. Kawamura, and T. Maruyama, "Void Formation in Magnetite Scale Formed on Iron at 823 K -Elucidation by Chemical Potential Distribution-," *Materials Science Forum*, vol. 522–523, pp. 37–44, Aug. 2006, doi: 10.4028/www.scientific.net/MSF.522-523.37.

[18] A. Kaderi, A. Z. Mohd Zainal, M. H. Ani, and R. Othman, "Observation On Void Formed In Oxide Scale Of Fe-Cr-Ni Alloy At 1073k In Dry And Humid Environments," *IIUM Engineering Journal*, vol. 12, no. 5, Jan. 2012, doi: 10.31436/iiumej.v12i5.235.

[19] M. E. G. Lyons and M. P. Brandon, "Redox switching and oxygen evolution electrocatalysis in polymeric iron oxyhydroxide films," *Physical Chemistry Chemical Physics*, vol. 11, no. 13, p. 2203, 2009, doi: 10.1039/b815338h.

[20] P. Bai and M. Z. Bazant, "Charge transfer kinetics at the solid–solid interface in porous electrodes," *Nat Commun*, vol. 5, no. 1, p. 3585, Apr. 2014, doi: 10.1038/ncomms4585.

[21] N. M. Marković, B. N. Grgur, and P. N. Ross, "Temperature-Dependent Hydrogen Electrochemistry on Platinum Low-Index Single-Crystal Surfaces in Acid Solutions," *J Phys Chem B*, vol. 101, no. 27, pp. 5405–5413, Jul. 1997, doi: 10.1021/jp970930d.

[22] T. T. Belete, M. C. M. Van De Sanden, and M. A. Gleeson, "Effects of transition metal dopants on the Calcination of CaCO3 under Ar, H2O and H2," *Journal of CO2 Utilization*, vol. 31, pp. 152–166, May 2019, doi: 10.1016/j.jcou.2019.03.006.

[23] F. Van Loock, V. Bernardo, M. A. Rodríguez Pérez, and N. A. Fleck, "The mechanics of solid-state nanofoaming," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 475, no. 2230, 2019, doi: 10.1098/rspa.2019.0339.

[24] M. Benitez-Guerrero, B. Sarrion, A. Perejon, P. E. Sanchez-Jimenez, L. A. Perez-Maqueda, and J. Manuel Valverde, "Large-scale high-temperature Solar Energy storage using Natural Minerals," *Solar Energy Materials and Solar Cells*, vol. 168, pp. 14–21, Aug. 2017, doi: 10.1016/j.solmat.2017.04.013.

[25] [Y. F. Gong and B. C. De Cooman, "Kirkendall void formation during selective oxidation," *Metall Mater Trans A Phys Metall Mater Sci*, vol. 41, no. 9, pp. 2180–2183, Sep. 2010, doi: 10.1007/s11661-010-0356-6.

https://doi.org/10.31436/iiumej.v26i3.3671

MACHINING PERFORMANCE WITH RESPECT TO CUTTING FORCES, VIBRATIONS AND SURFACE QUALITY IN DRILLING OF HYBRID ABACA AND GLASS FIBERS REINFORCED POLYMER COMPOSITE

MOHD. IQBAL[1*], TEUKU FIRSA[1], NAZARUDDIN[2], INDERA SAKTI NASUTION[3], MUHAMMAD RIZAL[1], JIHADUL QAHHAR[1], IRFAN HILMY[4]

[1]*Department of Mechanical and Industrial Engineering, Universitas Syiah Kuala, Banda Aceh, Indonesia*
[2]*Informatics Department, Faculty of Science, Universitas Syiah Kuala, Banda Aceh, Indonesia*
[3]*Agricultural Engineering Department, Faculty of Agriculture, Universitas Syiah Kuala, Banda Aceh, Indonesia*
[4]*Higher Colleges of Technology, Fujairah Hulaifat Campus, Uni Arab Emirates*

**Corresponding author: mohd.iqbal@usk.ac.id*

(Received: 3o March 2025; Accepted: 13 May 2025; Published online: 9 September 2025)

ABSTRACT: Initially, polymer composites utilized synthetic fibers as reinforcements due to their high strength and excellent water resistance. However, synthetic fibers are non-biodegradable, and their manufacturing process involves chemicals, rendering them environmentally unfriendly. Consequently, natural fibers began to be employed owing to their low production costs and inherent biodegradability. Nevertheless, natural fibers possess lower strength and non-uniform sizes, making them challenging to shape. Hybrid composites have been developed to address these limitations, combining synthetic and natural fibers to leverage their respective advantages and mitigate each other's shortcomings. As a result, hybrid composites are increasingly being adopted in various industries, including automotive, construction, sports, and electronics. During production, hybrid composite panels still require machining processes, such as milling to smooth the surface and drilling to create connecting holes, to become final products. Careful selection of machining parameters are needed to ensure product quality and minimize defects, including tool wear, fiber debonding, and delamination. Cutting forces and vibrations, critical factors influencing machining performance, can be regulated through optimized cutting conditions. This study examines the effects of cutting conditions on cutting forces and vibrations while drilling hybrid composites composed of abaca and glass fibers. Composite panels were fabricated using the compression molding method, consisting of 10% abaca fiber, 10% glass fiber, and 80% polyester resin with 1% hardener by weight. Drilling experiments were conducted using a two-flute high-speed steel (HSS) cutting tool under varying spindle speeds and feed rates. The results revealed that cutting parameters significantly influenced machining behaviour. Specifically, higher spindle speeds increased cutting forces by up to approximately 15%, whereas higher feed rates amplified vibration acceleration by up to 67.8%. These findings contribute to a deeper understanding of machining performance and provide valuable insights for optimizing drilling parameters to enhance machining efficiency and surface quality in hybrid natural–synthetic fiber composites.

ABSTRAK: Pada mulanya, komposit polimer menggunakan serat sintetik sebagai penguat kerana kekuatannya yang tinggi dan rintangan air terbaik. Walau bagaimanapun, serat sintetik tidak terbiodegradasi, dan proses pembuatannya melibatkan bahan kimia, menjadikannya tidak mesra alam. Akibatnya, serat semula jadi mula digunakan kerana kos pengeluarannya yang rendah dan kebolehbiodegradan. Namun begitu, serat semulajadi mempunyai kekuatan

https://doi.org/10.31436/iiumej.v26i3.3671

lebih rendah dan saiz tidak seragam, menjadikannya sukar dibentuk. Bagi menangani masalah ini, komposit hibrid dibangunkan, menggabungkan serat sintetik dan semulajadi bagi memanfaatkan kelebihan masing-masing dan mengurangkan kelemahan masing-masing. Hasilnya, komposit hibrid semakin diterima pakai dalam pelbagai industri, termasuk automotif, pembinaan, sukan dan elektronik. Semasa peringkat pengeluaran, panel komposit hibrid masih memerlukan proses pemesinan, seperti milling bagi melicinkan permukaan dan penggerudian bagi mencipta lubang penyambung, sebelum menjadi produk akhir. Pemilihan parameter pemesinan yang teliti diperlukan bagi memastikan kualiti produk dan meminimumkan kecacatan, termasuk haus alat, nyah ikatan gentian dan penyimpangan. Daya pemotongan dan getaran, yang merupakan faktor kritikal mempengaruhi prestasi pemesinan, dapat dikawal melalui keadaan pemotongan yang optimum. Kajian ini mengkaji kesan keadaan pemotongan ke atas daya pemotongan dan getaran semasa penggerudian komposit hibrid yang terdiri daripada serat abaka dan gentian kaca. Panel komposit telah difabrikasi menggunakan kaedah pengacuan mampatan, yang terdiri daripada 10% gentian abaka, 10% gentian kaca, dan 80% resin poliester dengan 1% pengeras mengikut berat. Eksperimen penggerudian telah dijalankan menggunakan alat pemotong keluli berkelajuan tinggi (HSS) dua mata di bawah kelajuan dan kadar suapan berbeza. Dapatan menunjukkan bahawa parameter pemotongan mempengaruhi tingkah laku pemesinan dengan ketara. Khususnya, kelajuan lebih tinggi menyebabkan daya pemotongan meningkat sehingga kira-kira 15%, manakala kadar suapan lebih tinggi menguatkan pecutan getaran sehingga 67.8%. Penemuan ini menyumbang kepada pemahaman yang lebih mendalam tentang prestasi pemesinan dan menyumbang pandangan berharga dalam mengoptimum parameter penggerudian bagi meningkatkan kecekapan pemesinan dan kualiti permukaan komposit serat semulajadi-sintetik hibrid.

KEYWORDS: *Hybrid Composite; Abaca and Glass Fiber; Drilling Process; Cutting Force; Vibration*

1. INTRODUCTION

The technical and economic advantages of Fiber Reinforced Polymer (FRP) composites stem from their properties differing from those of their individual constituent materials. The primary benefits of these composites include high strength, toughness, stiffness, low weight, and excellent creep resistance, which contribute to reduced risks of corrosion, wear, and fatigue compared to conventional materials. These characteristics highlight the remarkable versatility of FRP composites in meeting diverse industrial requirements across mechanical, electrical, magnetic, optical, and thermal properties, capabilities that monolithic materials cannot achieve. Consequently, these advantages make FRP composites highly desirable for various applications, from commercial aircraft development to sports equipment [1].

Fibers used as reinforcement in polymer composites are classified into two types: synthetic fibers and natural fibers. Synthetic fibers are produced through chemical processes using non-natural materials, typically derived from petrochemical polymers. These fibers exhibit excellent mechanical and physical properties; however, their manufacturing process generates chemical waste and may cause skin irritation in workers and users. Additionally, synthetic fibers are non-biodegradable, posing potential environmental concerns [1].

In contrast, natural fibers originate from renewable sources such as plants, animals, or minerals and are produced without synthetic chemicals or artificial processes. Commonly used natural fibers in industrial applications include cotton, hemp, jute, abaca, wool, silk, and cashmere. These fibers offer significant environmental advantages: they are renewable, biodegradable, and recyclable. However, natural fibers generally have lower resistance to extreme weather conditions, humidity, and UV radiation, making them more susceptible to

https://doi.org/10.31436/iiumej.v26i3.3671

degradation or mold growth under certain conditions. Furthermore, they tend to be less dimensionally stable, as they may shrink or stretch after washing [2].

Researchers have developed hybrid composites to address this issue by combining two types of fibers. Typically, one fiber in a hybrid composite has a lower modulus or cost, such as glass or Kevlar fiber. In comparison, other fibers, such as boron or carbon fiber, possess a higher modulus or cost. This combination optimizes the material's mechanical properties and cost-effectiveness. Consequently, natural fibers used as reinforcements in polymer composites can be combined with synthetic fibers, such as hemp and glass fibers [3], jute and Kevlar [4], or carbon and hemp [5].

Abaca fiber is a cellulose-based fiber obtained from the pseudostem of the abaca (Musa textilis) or banana plant. One of the key economic advantages of the abaca plant is its ability to grow without requiring fertilizers, pesticides, or large amounts of water [6]. Abaca fiber is widely recognized for its high strength, resistance to saltwater, and durability. Its long-term resilience and flexibility make it suitable for various applications. Additionally, abaca fiber is classified as a lightweight natural fiber, contributing to product strength without adding significant weight. Moreover, abaca fiber is environmentally friendly due to its biodegradability, which prevents environmental pollution. However, despite its advantages, abaca fiber has limitations in terms of thickness and length variation, which may result in unevenness and reduced resistance to sudden impacts or pressure [7].

Glass fiber is produced by drawing molten glass into thin threads that can be spun. It is widely utilized in various industries for manufacturing products such as automobile bodies, aircraft, and ships. Glass fiber exhibits high tensile strength, making it an excellent reinforcing material in composite structures. Despite its strength, it remains relatively lightweight, which helps reduce the overall weight of the final product. Additionally, glass fiber exhibits resistance to corrosion, chemicals, and moisture, which contributes to its durability. It also possesses excellent thermal and electrical insulation properties and can withstand high temperatures. The ability to be molded into various sizes and shapes, along with its recyclability, enhances its appeal as an environmentally friendly material. However, glass fiber has certain limitations. It is less impact-resistant and may crack or break when subjected to sudden force or pressure. Moreover, its rough texture can cause discomfort upon direct skin contact, limiting its application in consumer products that require frequent handling. The glass fiber manufacturing process is energy-intensive, which impacts both production costs and environmental sustainability. Compared to certain synthetic fibers, such as carbon fiber, glass fiber is more brittle and less resistant to fatigue or long-term structural degradation [8, 9].

Utilizing hybrid composites involves integrating two types of fibers within a single composite structure. A particularly advantageous approach is the combination of natural and synthetic fibers, as this can result in composite materials that are strong, cost-effective, and environmentally friendly. By incorporating abaca fibers and glass fibers, the composite's specific mechanical and physical properties can be enhanced, leading to improved product quality and a more efficient manufacturing process [10].

Raj et al. [11] investigated the mechanical properties of hybrid composites reinforced with abaca-glass and abaca-Kevlar fibers. The results revealed that incorporating glass and Kevlar fibers significantly enhanced the tensile, flexural, and impact strengths, rendering the hybrid composites superior to the abaca composite alone. Although the abaca-Kevlar hybrid composite exhibited greater stability, the abaca-glass hybrid composite proved more efficient, as its strength was only marginally lower than that of the glass composite alone. Prakash et al. [12] investigated the abrasive wear behavior of abaca-glass fiber hybrid composites using a

pin-on-disc tribo-test. They reported that the hybridization of glass fibers into abaca fibers and an epoxy matrix significantly enhanced the wear properties, making it a qualified high-wear-resistance polymer composite. Qahhar et al. [13] studied the vibration in the milling of abaca-glass fiber hybrid composites. They found that cutting parameters, such as spindle speed, feed rate, and depth of cut, significantly affected the vibration. The study concluded that vibration can be controlled by selecting appropriate cutting conditions, reducing the tool wear rate, and maintaining surface quality.

In general, composites are not final products; additional processing steps are required, such as cutting to achieve the desired shape and size and drilling to create holes for assembly purposes. However, drilling composite materials presents several challenges, including 'delamination', which occurs when composite layers separate during machining [14]; 'tool wear', which shortens tool lifespan and increases production costs [15]; and 'excessive heat generation', which can degrade the polymer matrix, cause thermal damage, and reduce the mechanical properties of the composite [16]. Additionally, sudden variations in cutting force can lead to process instability and unpredictability [17], while excessive vibration may result in rough surfaces, reduced precision, and lower cutting accuracy and quality [18, 19]. These factors can hinder the attainment of optimal machining results for composite materials.

To mitigate these issues, it is essential to optimize cutting conditions by selecting appropriate cutting parameters, using sharp and suitable cutting tools for composite materials, and ensuring adequate cooling to prevent excessive heat buildup. Proper techniques like cutting in the correct direction or using lubricants can help minimize tool wear and avoid delamination. Moreover, routine maintenance of cutting equipment plays a crucial role in maintaining cutting efficiency and accuracy. By applying these strategies, cutting quality can be enhanced, and potential machining issues can be minimized, ensuring that the final composite product remains high in quality and performance [20-22].

This study investigates the cutting force and vibration generated during drilling a hybrid composite composed of abaca fiber and glass fiber. The composite was fabricated using the press molding method, with polyester resin as the matrix and abaca and glass fibers as the reinforcing materials. Drilling was performed using a two-flute high-speed steel (HSS) cutting tool with a 12 mm diameter, while varying the cutting parameters, specifically spindle speed and feed rate.

2. RESEARCH METHODS

2.1.Preparation of the materials

The specimen mold was constructed from a steel plate featuring a square pocket measuring 200 mm on each side with a thickness of 5 mm. Abaca fibers, obtained from CV Naturi, Aceh, Indonesia, were combed to obtain a uniform maximum diameter of 1 mm, then rinsed with water and dried in the sun, and finally cut into uniform lengths of 20 mm. The glass fiber (chopped strand mat type 120E, procured from Chemposite Store, Surabaya, Indonesia) was cut into 200 mm × 200 mm squares. Each piece of glass fiber with this dimension weighed 10 grams. The fabrication of the hybrid abaca-glass fiber composite panel is illustrated in Figure 1, following these steps.

Polyester resin and hardener (Yukalac 157 BQTN-EX Series, manufactured by Justus, Indonesia) were mixed in a ratio of 100:1 (Figure 1a). A layer of wax was applied to the bottom surface of the mold to prevent the composite panel from adhering to the mold. The polyester resin was poured evenly into the mold (Figure 1b). Next, 10 grams of abaca fiber were

https://doi.org/10.31436/iiumej.v26i3.3671

randomly distributed within the mold (Figure 1c), followed by another uniform layer of polyester resin. Two layers of glass fiber, weighing 20 grams, were placed over the abaca fiber (Figure 1d), and another layer of polyester resin was poured uniformly onto the glass fiber. As the final layer, 10 grams of abaca fiber were randomly distributed over the glass fiber (Figure 1e), and a final coating of polyester resin was applied.

A steel plate was placed on top of the mold (Figure 1f), and the mold was then positioned in a press machine, where pressure was applied until the steel plate and mold were securely locked in place (Figure 1g). The mold remained closed for four hours, allowing sufficient time for the natural curing process. Once cured, the composite panel was removed from the mold and prepared for drilling (Figure 1h). The total weight of the composite panel was 201 grams, resulting in a fiber-to-matrix weight ratio of 20:80.

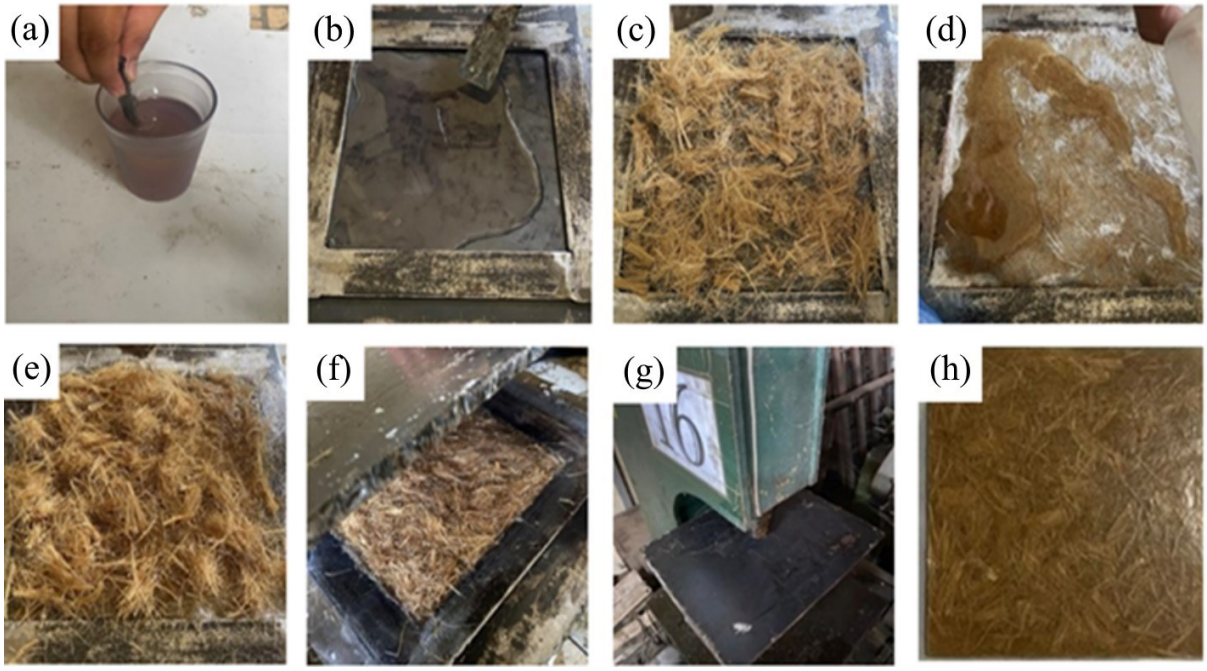


Figure 1. The Fabrication process of the composite panel: (a) mixing polyester resin with hardener, (b) pouring the polyester resin into the mold, (c) placing abaca fibers for the first layer, (d) placement of glass fiber, (e) placing abaca fibers for the final layer, (f) placement of the steel mold, (g) application of pressure load and (h) the composite panel.

2.2.Experimental Setup

The configuration of the drilling process and the equipment used to measure cutting force and acceleration is shown in the experimental setup (Figure 2). The hybrid abaca-glass fiber composite panel was securely clamped on top of the dynamometer (semiconductor stationary dynamometer with force sensing and piezoresistive strain sensors) using a jig and fixture [23]. An accelerometer sensor (PCB Piezotronics, model 352C33) was attached to the test object. Data acquisition was performed by connecting the system to a computer, where the Neo-MoMac software was used: National Instruments NI 9237 for recording the cutting force signal and National Instruments NI 9520 for capturing the vibration signal generated during the drilling process.

This research investigates the effect of cutting parameters on cutting force and vibration during the drilling of hybrid abaca-glass fiber composites. The cutting parameters considered in this study were spindle speed and feed rate, which are crucial in determining process

https://doi.org/10.31436/iiumej.v26i3.3671

efficiency and drilling quality. Variations in spindle speed and feed rate were applied to analyze their influence on cutting forces, cutting vibrations, drill hole surface quality, and overall drilling performance. Three levels of each parameter were defined based on the tool catalog recommendations for drilling composite materials using an HSS tool, as shown in Table 1.

The experiment was designed using a complete factorial configuration. The Design of Experiment (DOE) resulted in nine cutting conditions for two variables with three levels each. Five drillings were performed for each condition, and the average value was taken as the final measurement. The distances between the holes in the composite panel were 20 mm and 37 mm, resulting in 15 holes per panel. Three composite panels were prepared, each used to drill five holes for each of the nine cutting conditions.

The drilling was performed using an Agma A-8 CNC machine with a 12 mm, 2-flute HSS drill tool (Nachi, type L/N 500, procured from Sumber Teknik Mandiri, Jakarta, Indonesia). Cutting force and vibration signals were recorded from when the cutting tool made contact with the composite surface until it retracted after penetration, and the cutting process stopped. The measurement results were displayed graphically on the computer screen and exported to a Microsoft Excel file.

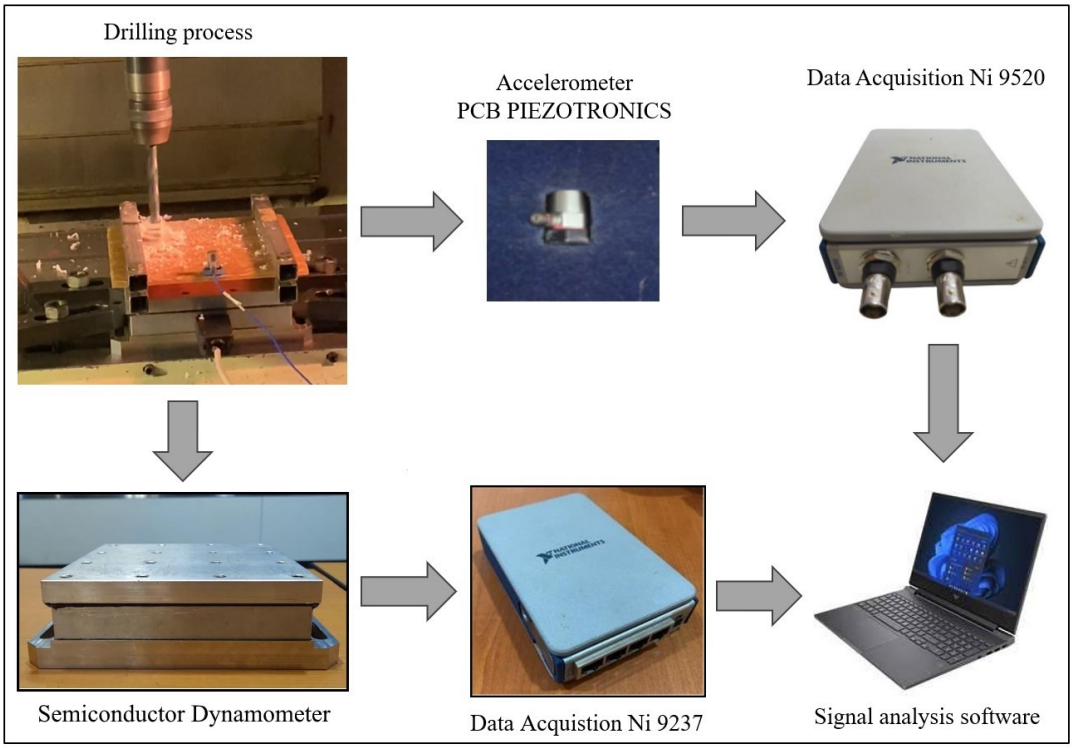


Figure 2. The experimental setup.

Table 1. Experimental parameters

No.	Level	Spindle Speed, N (rpm)	Feed Rate, F (mm/rev)
1	-1	1000	0.10
2	0	1500	0.15
3	1	2000	0.20

https://doi.org/10.31436/iiumej.v26i3.3671

3. RESULTS AND DISCUSSION

3.1.The drilling process and data collection

Figure 3 shows the abaca-glass fiber hybrid composite panels after the drilling process. During the drilling process, the cutting forces were measured using the dynamometer and recorded in volts by the NI 9237 data acquisition system. The recorded data were then converted to Newtons by multiplying them by the corresponding sensitivity values. The sensitivity values were 2.365 for the x-axis, 2.279 for the y-axis, and 3.806 for the z-axis. The cutting force signals (in Newtons) were then plotted using MATLAB, as shown in Figure 4.

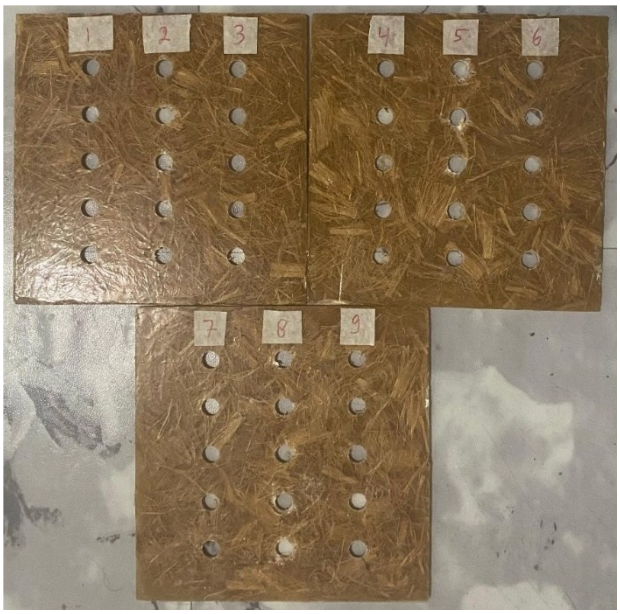


Figure 3. The abaca-glass fiber hybrid composite panels after the drilling process.

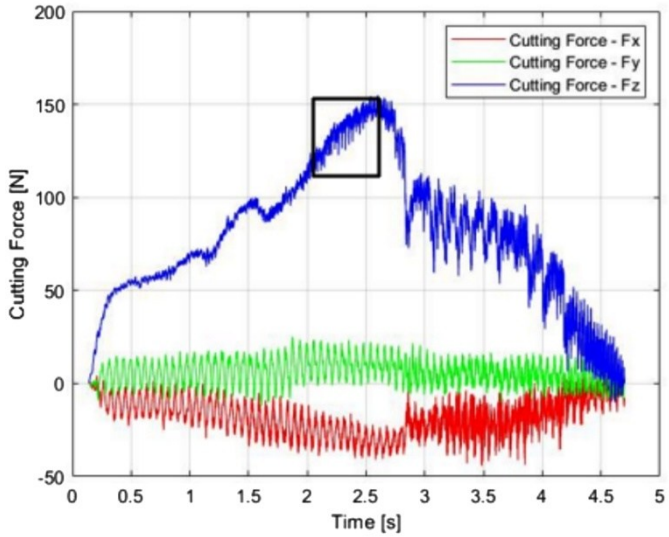


Figure 4. The cutting force signals.

https://doi.org/10.31436/iiumej.v26i3.3671

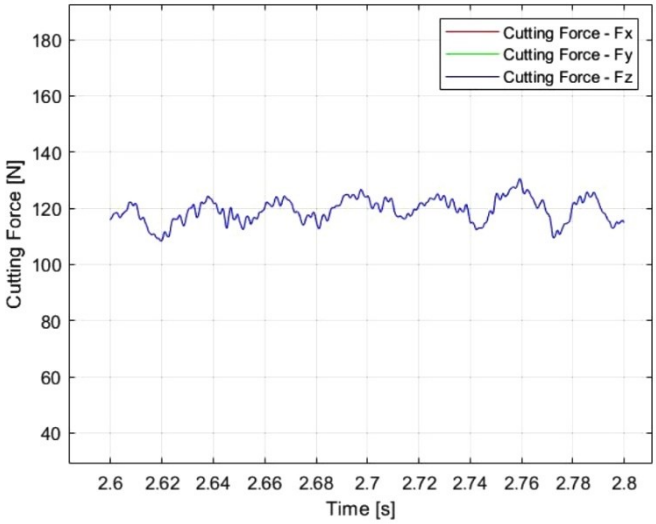


Figure 5. Cutting force signals after being processed.

As shown in Figure 4, the dominant cutting force was Fz (represented by the blue signal), which corresponds to the z-axis direction, or the axial/vertical direction of the cutting tool. The cutting force signal appeared when the drill bit first made contact with the surface of the composite panel and increased as the tool continued to penetrate the material. When the entire cutting edge of the drill was in complete contact with the composite surface, the force signal reached its maximum level. As the drill bit exited the opposite side of the composite panel, the force signal gradually decreased. The cutting force value was determined by averaging the uniform signal at the maximum level, as shown in Figure 5.

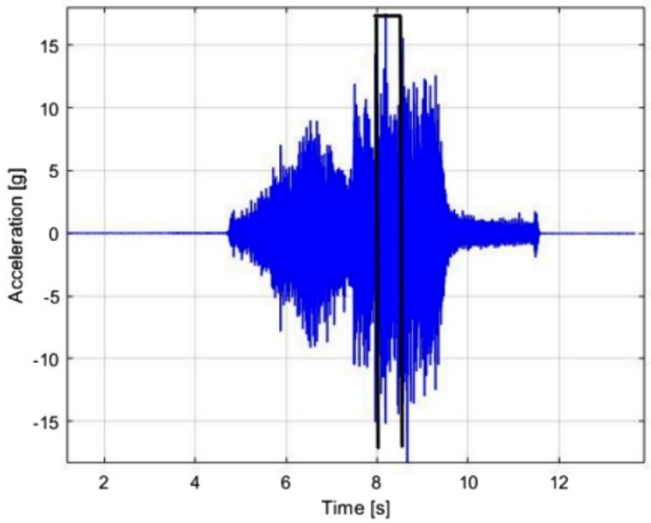


Figure 6. The recorded vibration signal.

The vibration signal was recorded from the moment the drill bit made contact with the surface of the composite panel until it exited the drilled hole, as shown in Figure 6. The maximum vibration signal occurred when the drill bit was in complete contact with the inner surface of the hole in the composite panel. The vibration data were obtained by averaging the signal at this maximum level.

https://doi.org/10.31436/iiumej.v26i3.3671

After drilling, the vibration signal was measured to evaluate changes resulting from the interaction between the drill bit and the composite panel. This measurement aimed to analyze the vibration acceleration level after drilling and identify patterns that might indicate imbalances or irregularities in the process. The obtained data were processed using MATLAB to visualize the vibration characteristics more clearly. Figure 7 presents the recorded vibration signal after the drilling process.

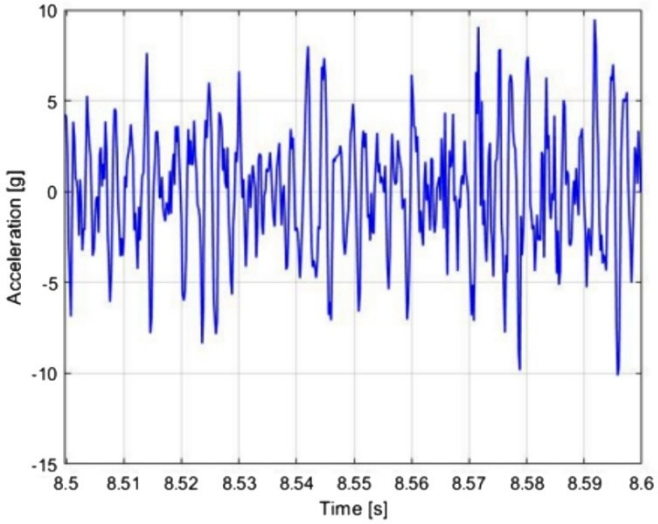


Figure 7. The vibration signal after processing.

Table 2. Result of the experiment

	Diameter tools		12 mm	
No.	**Spindle Speed N (rpm)**	**Feed Rate F (mm/rev)**	**Cutting Force (Fz, N)**	**Acceleration (m/s^2)**
1	1000	0.1	154.60	0.028
2	1000	0.15	191.67	0.047
3	1000	0.2	233.95	0.038
4	1500	0.1	151.62	0.032
5	1500	0.15	161.97	0.05
6	1500	0.2	224.33	0.041
7	2000	0.1	187.85	0.04
8	2000	0.15	191.75	0.06
9	2000	0.2	248.51	0.054

The experimental results are presented in Table 2. Each cutting condition was tested five times, measuring cutting force and vibration simultaneously. The final result was obtained by averaging the five recorded values.

3.2. The Effect of Spindle Speed on Cutting Force

Figure 8 illustrates the effect of spindle rotation speed on cutting force at different feed rates. When the spindle speed increased from 1000 to 1500 rpm, the cutting force decreased across all feed rate levels. However, the reduction was minimal at feed rates of 0.1 mm/rev and 0.2 mm/rev. In contrast, at a feed rate of 0.15 mm/rev, there was a significant decrease of 18.6%, from 191.67 N to 161.97 N. Conversely, when the spindle speed increased from 1500

https://doi.org/10.31436/iiumej.v26i3.3671

rpm to 2000 rpm, the cutting force increased uniformly across all feed rate levels by approximately 15%.

As shown in Figure 8, spindle speed has a significant impact on cutting force. In many drilling operations, higher spindle speeds produce smaller chip sizes, reducing the cutting force required for material removal. However, increased spindle speed generates more heat due to friction between the cutting tool and the workpiece. Excessive heat can accelerate tool wear, alter the cutting geometry, and increase the cutting force.

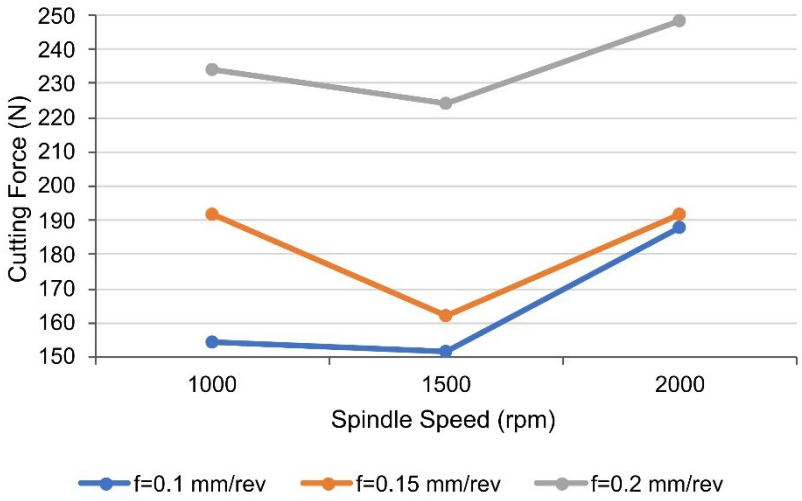


Figure 8. The Effect of Spindle Speed on Cutting Force.

3.3. The Effect of Feed Rate on Cutting Force

The impact of feed rate on cutting force at different spindle speeds is shown in Figure 9. The cutting force increased at all spindle speed levels when the feed rate was raised from 0.1 mm/rev to 0.15 mm/rev. However, the increase was negligible at spindle speeds of 1500 rpm and 2000 rpm. At a spindle speed of 1000 rpm, there was a significant rise of 24%, from 154.60 N to 191.67 N. When the feed rate was further increased from 0.15 mm/rev to 0.2 mm/rev, the cutting force increased sharply across all spindle speed levels, reaching up to 39%.

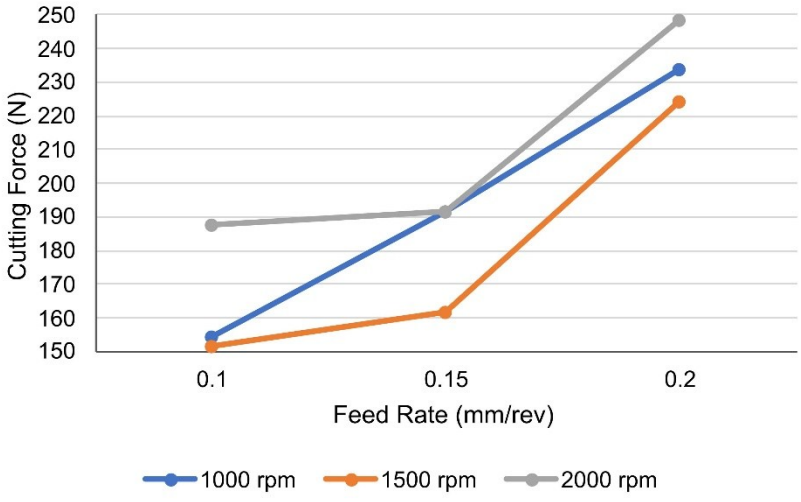


Figure 9. The Effect of Feed Rate on Cutting Force.

https://doi.org/10.31436/iiumej.v26i3.3671

Figure 9 demonstrates that cutting force is directly proportional to the feed rate. As the feed rate increases, the drill bit engages with a larger volume of material during each rotation, requiring greater force to cut through the material. At higher feed rates, the cutting edge of the drill bit encounters increased resistance due to the greater material removal per pass, leading to higher thrust and radial forces.

3.4. The Effect of Spindle Speed on Acceleration Vibration

Figure 10 illustrates the effect of spindle speed on acceleration vibration at different feed rates. As the spindle speed increased from 100 rpm to 1500 rpm and then to 2000 rpm, the acceleration increased slightly across all feed rate levels. However, the increase was more pronounced when the spindle speed rose from 1500 rpm to 2000 rpm compared to the increment from 1000 rpm to 1500 rpm.

The graph in Figure 10 demonstrates that spindle speed plays a crucial role in influencing vibrations. An increase in spindle speed leads to higher vibration frequency and amplitude at certain speeds due to resonance effects. As observed in previous results, higher spindle speeds generate larger cutting forces, which may increase vibration frequency and amplitude.

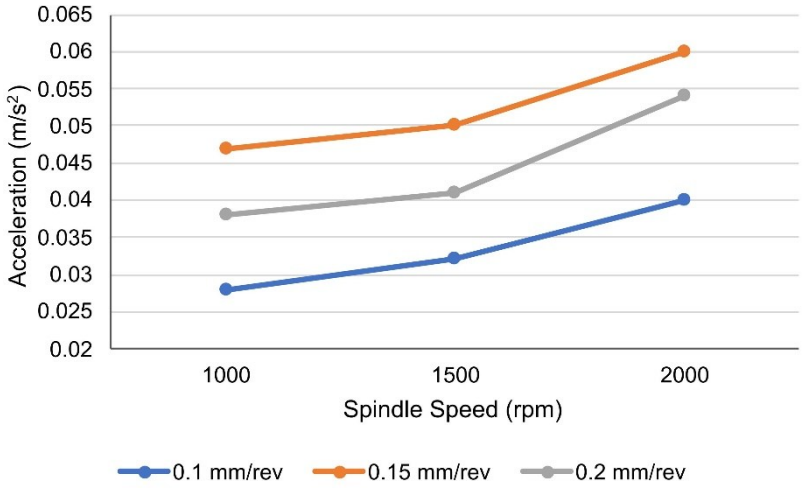


Figure 10. The Effect of spindle speed on acceleration.

3.5. The Effect of Feed Rate on Acceleration Vibration

The impact of feed rate on acceleration vibration at varying spindle speeds is illustrated in Figure 11. The acceleration increased sharply at each spindle speed level when the feed rate was raised from 0.1 to 0.15 mm/rev. A maximum increase of 67.8% was observed at a spindle speed of 1000 rpm, followed by 56.25% at 1500 rpm and 50% at 2000 rpm. However, when the feed rate increased from 0.15 mm/rev to 0.2 mm/rev, the acceleration decreased across all spindle speed levels. The reduction reached 19% at 1000 rpm, 18% at 1500 rpm, and 10% at 2000 rpm.

The results indicate that the feed rate significantly influences vibration during the drilling of hybrid composite panels. However, the relationship is not linear. In this case, lower acceleration vibration can be achieved by applying either a low feed rate (0.1 mm/rev) or a high feed rate (0.2 mm/rev).

https://doi.org/10.31436/iiumej.v26i3.3671

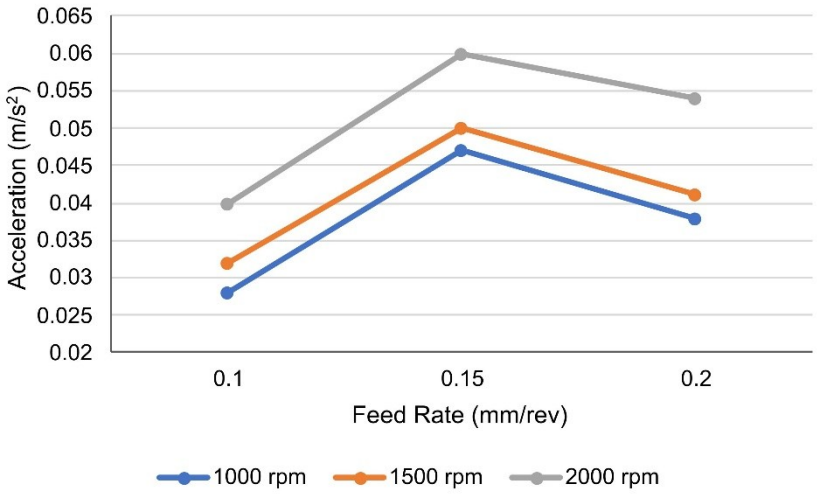


Figure 11. The Effect of Feed Rate on Acceleration.

3.6. The Influence of Hole Drilling Parameters on Hole Quality

Figure 12 shows the quality of holes in each test based on the hole quality images from the drilling process. The hole quality was assessed through visual inspection, either directly on the hole surface or from images captured by a digital camera. No measuring equipment was used in this evaluation. A hole with a smooth surface was considered good quality, whereas a hole with a rough surface was classified as poor quality.

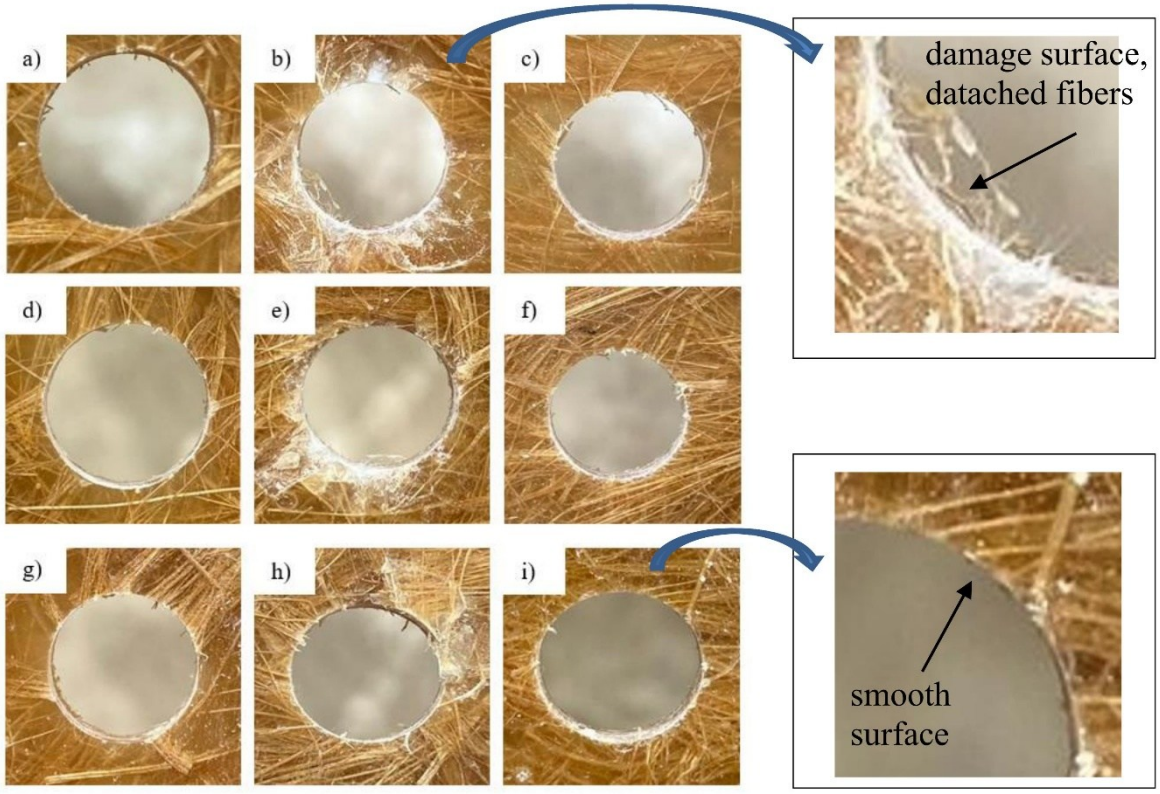


Figure 12. Data on cutting results related to hole quality: spindle speed at 1000 rpm (a) f=0.1 mm/rev, (b) f=0.15 mm/rev, (c) f=0.2 mm/rev, spindle speed at 1500 rpm (d) f=0.1 mm/rev, (e) f=0.15 mm/rev, (f) f=0.2 mm/rev, and spindle speed at 2000 rpm (g) f=0.1 mm/rev, (h) f=0.15 mm/rev, (i) f=0.2 mm/rev.

Figure 12 shows that increased spindle rotation speed results in better hole quality. The best hole quality was achieved at a spindle speed of 2000 RPM with a feed rate of 0.2 mm/rev, as shown in Figure 12.(i), where the hole surface appeared smooth and uniform. In contrast, the poorest hole quality was observed at a spindle speed of 1000 RPM with a feed rate of 0.15 mm/rev, as shown in Figure 12.(b), where some parts of the surface were damaged, and some fibers were detached from the panel.

Higher spindle speeds can increase cutting efficiency, reducing the fiber pull-out and matrix cracking, which can cause rough surfaces. Additionally, the cutting tool can cut fibers more effectively, reducing fiber breakage and resulting in a smoother surface. Furthermore, the cutting tool can penetrate the composite panel more smoothly, reducing the material layer separation risk and minimizing delamination [14, 15]. However, it is worth noting that excessively high spindle speeds can generate excessive heat, leading to matrix degradation and tool wear. Therefore, optimizing spindle speed is crucial to achieving a smooth surface when drilling composite panels.

4. CONCLUSION

The research findings indicate that higher spindle speed and feed rate result in increased cutting force and vibration levels, with the maximum cutting force of 248.515 N observed at 2000 RPM and 0.2 mm/rev, and the minimum cutting force of 151.619 N recorded at 1500 RPM and 0.1 mm/rev. Similarly, vibration levels peaked at 0.06 m/s² at 2000 RPM and 0.15 mm/rev, while the lowest vibration level of 0.028 m/s² was recorded at 1000 RPM and 0.1 mm/rev. Cutting conditions also affected hole quality, where the best results were achieved at 2000 RPM and 0.2 mm/rev, and the poorest quality at 1000 RPM and 0.15 mm/rev. These outcomes contribute valuable insights to composite machining and can guide the industry toward more efficient and higher-quality drilling practices. However, this study was limited to cutting force, vibration, and surface quality, and future work should investigate tool wear, as some conditions that reduce cutting force and vibration while improving surface finish may also accelerate tool degradation. Moreover, optimization studies are recommended to establish cutting conditions that balance surface quality, cost, and machining time.

ACKNOWLEDGEMENT

The authors would like to thank Universitas Syiah Kuala for funding the research that made this work possible.

REFERENCES

[1] Hasan, M., Zhao, J., & Jiang, Z. (2019). Micromanufacturing of composite materials: a review. International Journal of Extreme Manufacturing, 1(1), 012004.

[2] Madhu, P., Sanjay, M. R., Senthamaraikannan, P., Pradeep, S., Saravanakumar, S. S., & Yogesha, B. (2019). A review on synthesis and characterization of commercially available natural fibers: Part-I. Journal of Natural Fibers, 16(8), 1132–1144.

[3] Gupta, A., Shohel, S. M., Singh, M., & Singh, J. (2024). Study on mechanical properties of natural fiber (Jute)/synthetic fiber (Glass) reinforced polymer hybrid composite by representative volume element using finite element analysis: A numerical approach and validated by experiment. Hybrid Advances, 6, 100239.

[4] Bhuiyan, A. A. H., Hossain, M. F., Rana, M. S., & Ferdous, M. S. (2023). Impact of fiber orientations, stacking sequences and ageing on mechanical properties of woven jute-kevlar hybrid composites. Results in Materials, 20, 100477.

[5] Hassan, S. A., Binoj, J. S., Goh, K. L., Mansingh, B. B., Varaprasad, K. C., Yahya, M. Y., ... & Wulandari, A. P. (2022). Effect of fiber stacking sequence and orientation on quasi-static indentation properties of sustainable hybrid carbon/ramie fiber epoxy composites. Current Research in Green and Sustainable Chemistry, 5, 100284.

[6] Iqbal, M., Aminanda, Y., Firsa, T., Nazaruddin, Nasution, I. S., Erawan, D. F., ... & Nasution, A. R. (2023, January). The effect of fiber content and fiber orientation on bending strength of abaca fiber reinforce polymer composite fabricated by press method. In AIP Conference Proceedings (Vol. 2643, No. 1, p. 050055). AIP Publishing LLC.

[7] Iqbal, M., Firsa, T., Nazaruddin, N., Nasution, I. S., Zulfadhli, Z., Aminanda, Y., ... & Pradana, A. (2023). Tensile Strength of Abaca Fiber Reinforced Polymer Composite Fabricated by Press Method: Effect of Fiber Content and Fiber Orientation. Key Engineering Materials, 951, 73-83.

[8] Nazaruddin, N., Akram, A., Hasanuddin, I., Iqbal, M., Kurniawan, R., & Putra, R. (2019, May). Mechanical properties of glass fiber reinforced polyester resin for use as the wall of the Acehnese boat 'Thep-Thep'. In IOP Conference Series: Materials Science and Engineering (Vol. 523, No. 1, p. 012080). IOP Publishing.

[9] Rizal, M., Mubarak, A. Z., & Alhaidar, T. R. (2024). A Comparative Study of the Static and Dynamic Characteristics of Jute and Glass Fiber-Reinforced Polyester Composites. Experimental Techniques, 48(2), 229-237.

[10] Dong, C. (2018). Review of natural fibre-reinforced hybrid composites. Journal of Reinforced Plastics and Composites, 37(5), 331-348.

[11] Raj, M. K. A., Muthusamy, S., Panchal, H., Ibrahim, A. M. M., Alsoufi, M. S., & Elsheikh, A. H. (2022). Investigation of mechanical properties of dual-fiber reinforcement in polymer composite. journal of materials research and technology, 18, 3908-3915.

[12] Prakash, V., Pradhan, S., Majhi, S., & Acharya, S. K. (2023). Evaluation of dry sliding wear behavior of abaca/glass fiber hybrid composite for tribological applications. Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials: Design and Applications, 237(3), 714-723.

[13] Qahhar, J., Iqbal, M., Mubarak, A. Z., & Abhang, L. B. (2025, May). The Influence of Cutting Conditions on Vibration in the Milling Process of Abaca-Glass Fiber Hybrid Composite. In Materials Science Forum (Vol. 1149, pp. 55-64). Trans Tech Publications Ltd.

[14] Babu, J., Sunny, T., Paul, N. A., Mohan, K. P., Philip, J., & Davim, J. P. (2016). Assessment of delamination in composite materials: a review. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 230(11), 1990-2003.

[15] Hakami, F., Pramanik, A., & Basak, A. K. (2017). Tool wear and surface quality of metal matrix composites due to machining: A review. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 231(5), 739-752.

[16] Esposito, L., Sorrentino, L., Penta, F., & Bellini, C. (2016). Effect of curing overheating on interlaminar shear strength and its modelling in thick FRP laminates. The International Journal of Advanced Manufacturing Technology, 87, 2213-2220.

[17] Islam, S., Yuan, S., & Li, Z. (2020). A cutting force prediction model, experimental studies, and optimization of cutting parameters for rotary ultrasonic face milling of C/SiC composites. Applied Composite Materials, 27, 407-431.

[18] Abhang, L. B., Iqbal, M., & Hameedullah, M. (2021, June). An Experimental Model for the Prediction of Chip Thickness in Steel Turning. In Proceedings of the 2nd International Conference on Experimental and Computational Mechanics in Engineering: ICECME 2020, Banda Aceh, October 13–14 (pp. 137-149). Singapore: Springer Singapore.

[19] Ilham, M. M., & Mufarrih, A. (2018). Effect of Point Angle on Delamination in Drilling of KFRP Composite. Jurnal Teknik Mesin, 8(2), 51-55.

[20] Alayaini, M., Rizal, M., Iqbal, M., & Dirhamsyah, M. (2022, September). Experimental Study of MQL Performance on the Titanium Material Drilling Process. In Materials Science Forum (Vol. 1067, pp. 35-45). Trans Tech Publications Ltd.

[21] Djafar, Z., Azis, N., & Zulkifli, T. (2023, July). Ramie fiber woven composite: the effect of feedrate variation on the tensile strength of the open hole in the drilling. In Materials Science Forum (Vol. 1092, pp. 27-34). Trans Tech Publications Ltd.

[22] Ikbal, M., Rizal, M., Ali, N., & Putra, T. E. (2023). Effect of Hybridization of Ramie Fibers on the Vibrational and Damping Responses of Ramie/Glass/Epoxy Resin Composite Laminates. Key Engineering Materials, 951, 65-71.

[23] Rizal, M., Ghani, J. A., Usman, H., Dirhamsyah, M., & Mubarak, A. Z. (2024). Development and testing of a stationary dynamometer using cross-beam-type force-sensing elements for three-axis cutting force measurement in milling operations. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 238(6-7), 950-961.

https://doi.org/10.31436/iiumej.v26i3.3497

DYNAMIC ANALYSIS ON AN AERIAL WORK PLATFORM USING A HYBRID CAD APPROACH FOR SATELLITE TESTING APPLICATIONS

NOR MOHD HAZIQ NORSAHPERI [1], SALMIAH AHMAD[2*], SITI FAUZIAH TOHA[3], MOHD AZRI ABD MUTALIB[4], MOHAMAD RIDZWAN ISHAK[5], DONG EUI CHANG[5]

[1]Department of Electrical and Electronics Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43000, UPM Serdang, Selangor, Malaysia
[2]College of Engineering and Technology, University of Doha for Science and Technology, Doha, 24449, Qatar
[3]Department of Mechatronics Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, 50728, Kuala Lumpur, Malaysia
[4]Machine Design Section, Machinery Technology Centre, SIRIM Berhad, 44200, Hulu Selangor, Selangor, Malaysia
[5]Department of Aerospace Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43000, UPM Serdang, Selangor, Malaysia
[6]School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

**Corresponding author: salmiah.ahmad@udst.edu.qa*

(Received: 11 December 2024; Accepted: 22 March 2025; Published online: 9 September 2025)

ABSTRACT: This paper presents a dynamic analysis using a hybrid CAD model approach, focusing on a motorized adjustable vertical platform's lifting and baseplate loading subsystems as a case study. The hybrid CAD model approach allows for examining mechanism behaviours such as acceleration torque, speed variation, and jerk rates, as well as their impact on the total torque required by the subsystems. The analysis starts by developing 3D and 2D models with velocity profiles, followed by comprehensive analyses. The most interesting part of the study is that while the acceleration torque rate is high, its influence is minimal due to small accelerations and decelerations at approximately -5×10^{-5} m/s^3 to 3.1×10^{-5} m/s^3 within the subsystem. Additionally, torque requirements for the lifting subsystem remain consistent at 631.22 N·m at the highest position and 364.16 N·m at the lowest position across different speed modes. The study also evaluates jerk rates during acceleration and deceleration to ensure compliance with ISO standards for ride quality. This approach shows promise for developing heavy-duty autonomous aerial work platforms, especially in the space industry, where understanding system behaviours is crucial before the development process begins.

ABSTRAK: Kajian ini membentangkan analisis dinamik menggunakan pendekatan model hibrid CAD, dengan tumpuan kepada subsistem pengangkatan dan pemuatan plat asas (baseplate loading subsystem) bagi platfom menegak boleh laras bermotor (motorised adjustable vertical platform) sebagai kajian kes. Pendekatan model hibrid CAD ini membolehkan pemerhatian terhadap tingkah laku mekanisme seperti kadar tork pecutan, variasi kelajuan, dan kadar kejutan, serta impaknya terhadap jumlah tork yang diperlukan oleh subsistem. Analisis dimulakan dengan pembangunan model 3D dan 2D dengan profil kelajuan, diikuti dengan analisis komprehensif. Penemuan paling menarik yang didapati pada kajian ini adalah walaupun kadar tork pecutan tinggi, pengaruhnya sangat minimum kerana pecutan dan penghentian kecil dalam subsistem iaitu sekitar -5 x 10-5 m/s3 hingga 3.1 x 10-5 m/s3. Tambahan, keperluan tork bagi mengangkat subsistem kekal konsisten pada 631.22 N.m pada posisi tertinggi dan posisi terendah pada 364.16 N.m merentasi pelbagai mod

kelajuan. Kajian ini juga menilai kadar kejutan semasa pecutan dan pengurangan kelajuan bagi memastikan pematuhan piawaian ISO bagi kualiti tunggangan. Pendekatan ini menunjukkan potensi dalam membangunkan platform kerja berat di udara secara autonomi, terutama dalam industri angkasa, di mana pemahaman tingkah laku sistem adalah penting sebelum proses pembangunan bermula.

KEYWORDS: *Dynamic Analysis, Scissor Mechanism, Satellite Test, and CAD Modelling*

1. INTRODUCTION

The thermal and vacuum test is one of the significant satellite tests that can be performed using a Thermal Vacuum Chamber (TVC), located in a clean room, as illustrated in Figure 1. To avoid any malfunction of the components in the satellite before the tests, as various components and the satellite itself are very sensitive to external particles, the cleanliness of the environment and specifications of the clean room are very important and shall be preserved [1, 2]. Commonly, the available facility for unloading and loading is the overhead crane, which poses a risk to the safety of the satellite and the operators because such a crane is sensitive to external force disturbances and hence, prone to oscillate [1, 3–5]. These limitations of the overhead crane significantly impact the operation and satellite safety, as the clean room is a confined space. Besides, the overhead crane is incongruous when loading and unloading, particularly when transferring the load inside the TVC. This is because the overhead crane does not have telescopic or horizontal motion, requires the travel system or rails to move, and has a limited working area. Hence, a customised adjustable vertical platform (MAVeP) is introduced to overcome the limitations of the overhead crane.

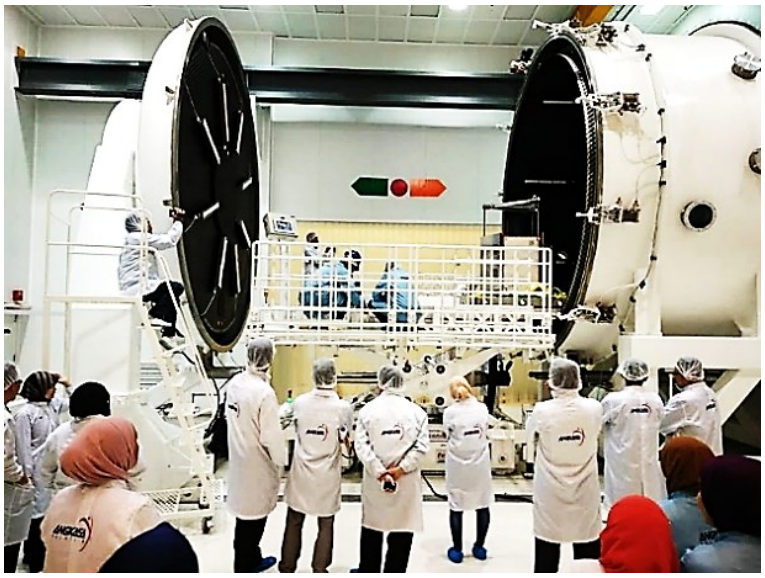


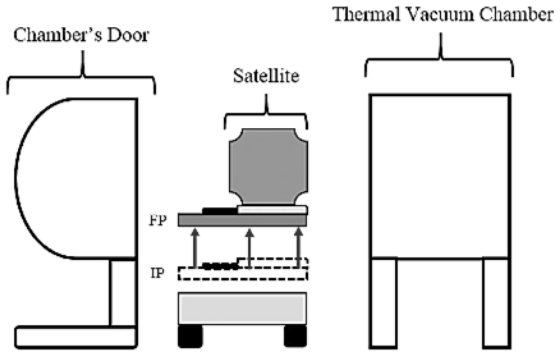
Figure 1. Thermal vacuum chamber overview.

Various analyses have been done to assess the mechanical forces acting on the actuator of an aerial work platform, such as the studies in [6–8] where most of these studies were carried out by conducting static analysis using the structural analysis method. Many other researchers have analysed the structure of the hydraulic actuator instead of the electrical-type actuator. For example, there is research on structural rigidity in [9], structure stiffness in [10], ground impact force in [11], platform stability in [12, 13] and hydraulic actuators in [14–17]. However, the only mechanical analysis done on the actuator is the static analysis using the structural analysis method. Not much research has been conducted on the dynamic analysis of the actuator, especially for the electrical-type actuator, compared to the hydraulic.

https://doi.org/10.31436/iiumej.v26i3.3497

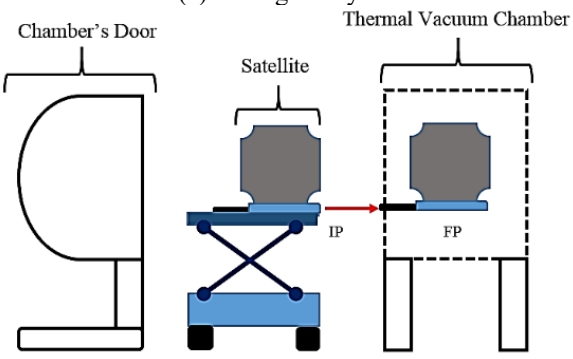
Several attempts have been made to conduct static and dynamic analyses of structures. The study in [18] proposes a higher-order strain-based element to enhance the accuracy of finite element analysis for plane structures by using higher-order polynomials to represent the deformation field. The strain-based approach has also been adopted in [19] to analyse the behaviour of materials near cracked structures. Other than the strain-based methods, the study in [20] introduces a dynamic model to analyse the behaviour of the aerial work platform during horizontal motion by considering mass distribution, external forces, and the system's dynamic response. Furthermore, the dynamic analysis was also conducted in [21] to evaluate the behaviour of the lifting arm system by developing the dynamic model of the system using the Lagrange equation. However, a significant problem with the aforementioned numerical approaches to studying the behaviour of a complex system is that they are challenging and prone to error.

Recent studies in [22–25] have suggested that computer-aided design (CAD) is becoming essential in analysing the dynamic behaviours of complex mechanical systems in modern engineering tools. In [26], the CAD model is adopted to enhance the accuracy of the reverse engineering process. The study in [27] also shows the importance of CAD techniques in obtaining a detailed and accurate model representing historical machinery. This allows for in-depth analysis of the equipment's design, construction, and functioning, providing insights into technological advancements and engineering practices. Other than the studies above, the CAD model has also been utilised to generate the robot path trajectories by using the mesh, as discussed in [28]. Despite ongoing efforts, it remains a continual challenge to work with highly complex surfaces using CAD techniques [29]. Moreover, there has been minimal discussion on the mechanical forces acting on the electrical actuators of an aerial work platform.



*FP: Final Position, IP: Initial Position

(a) Lifting Subsytem



*FP: Final Position, IP: Initial Position

(b) Baseplate Loading Subsystem

Figure 2. Vertical and horizontal positioning mechanisms

https://doi.org/10.31436/iiumej.v26i3.3497

Considering all, the primary aim of this paper is to conduct the dynamic analysis of the lifting and baseplate loading subsystems in MAVeP that can perform horizontal and vertical movements, as described in Figure 2 using the hybrid CAD model approach and considering the findings in [24]. The proposed hybrid CAD model approach utilizes both 3D and 2D models of MAVeP to conduct the dynamic analysis, as it has been reported that virtual models are one of the most practical tools for assisting practitioners in decision-making [29]. Additionally, this study aims to create a CAD model that closely resembles MAVeP, enabling a comprehensive illustration of the engineering significance of the invention and facilitating a deeper understanding of its functionality. To the best of the authors' knowledge, this study represents one of the earliest developments of an autonomous mobile aerial platform for satellite test facilities and clean room operations. Therefore, this study contributes significantly to research on dynamic analysis using the proposed hybrid CAD model approach and to the field of autonomous aerial work platforms for satellite test facilities.

Note that MAVeP is actuated by electric motors and functions within clean rooms or confined spaces as an additional satellite handling platform, particularly suited for thermal and vacuum testing. Moreover, the lifting and baseplate loading subsystems denote the mechanisms that provide vertical and horizontal motions, respectively. Several crucial analyses are required before selecting the electrical motor specifications and fabrication processes for MAVeP. These include designing the velocity profile, conducting acceleration and jerk analyses, and torque analysis for both the lifting and baseplate loading subsystems, using the actual specifications of MAVeP. The overall process to analyse the dynamic behaviour of the lifting and baseplate loading subsystems is shown in Figure 3.

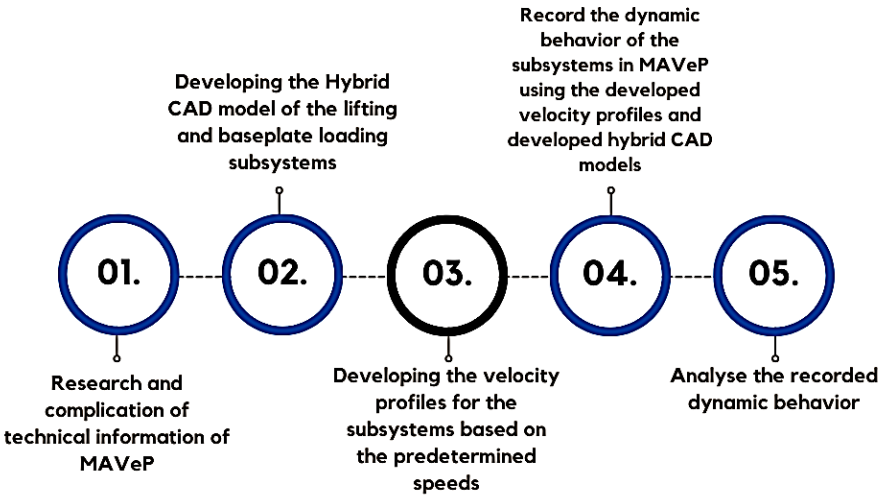


Figure 3**.** Flowchart of the dynamic analysis.

This paper is organized as follows: Section 2 describes the subsystem mechanisms and their properties, followed by Section 3, which explains the velocity profile design based on the prescribed subsystem speeds. The final two sections, 4 and 5, summarize the main findings of this study and present the conclusion, respectively.

2. HYBRID CAD MODELS

This section describes the mechanisms in MAVeP responsible for translating horizontal and vertical motions, referred to as the baseplate loading and lifting subsystems, respectively, along with their structural properties. The lifting subsystem utilizes a scissor mechanism, ball screw, and electrical motor, while the baseplate loading subsystem integrates a rack and pinion

https://doi.org/10.31436/iiumej.v26i3.3497

mechanism with an electrical motor. Then, this section explains the proposed hybrid CAD models of both subsystems for the dynamic analysis.

2.1. Vertical motion mechanism for lifting subsystem

The overall structural design of the MAVeP was designed, as shown in Figure 4, in the SOLIDWORKS 3D environment for the motion study and dynamic analysis. The structure consists of a multiple-link combination as depicted in Figure 5. At the stowed position, the minimum height of the platform is estimated to be 0.79 m, and the highest position the platform can reach is 1.48 m. The electrical motor was decided to be used for the mechanism, and the servomotor was selected. The servomotor for lifting the subsystem ($M1$) is fixed with a clamper ($C1$), as shown in Figure 5. The motor provides rotational motion, and the motion is converted into linear motion by a ball screw (BS) where the ball screw is mounted using a clamper ($C2$). Each clamper moves in linear motion and is attached with four links ($SS1$, $SS2$, $SS3$ and $SS4$). The end of each link ($SS1$, $SS2$, $SS3$ and $SS4$) is attached to the four main scissors (MS). If the power is fed to the motor, the motor shaft will rotate along with the ball screw, where the rotational motion is converted into linear motion, as in Figure 6. When the clampers ($C1$ and $C2$) is moving, each link ($SS1$, $SS2$, $SS3$ and $SS4$) rotates with respect to the rotational point on each clamper.

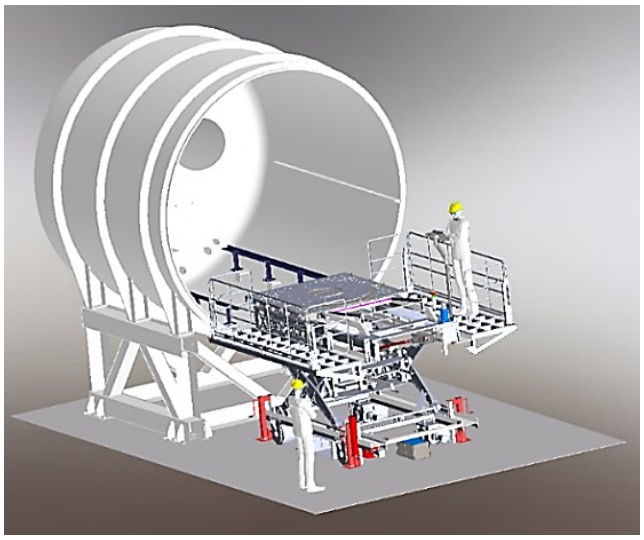


Figure 4. The MAVeP design in the SOLIDWORKS software.

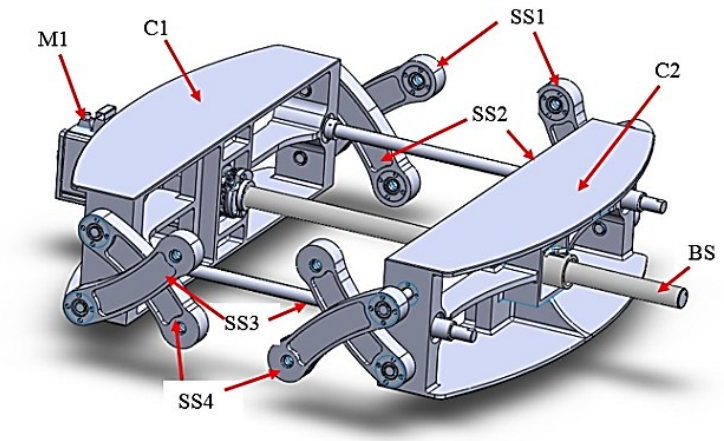


Figure 5. Clamper structure with the motor and ball Screw.

https://doi.org/10.31436/iiumej.v26i3.3497

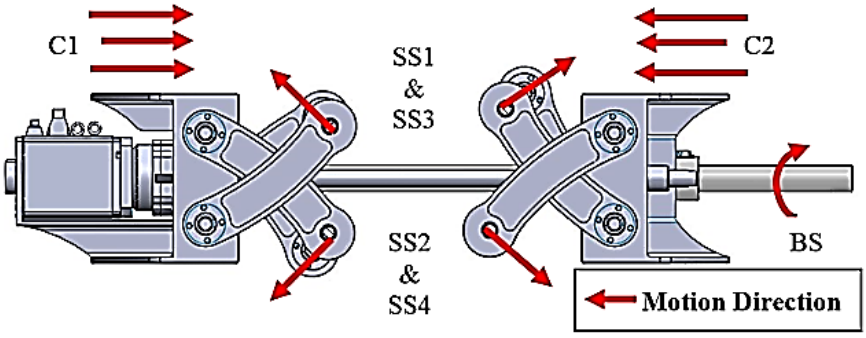


Figure 6. Clamper motion when the platform height is increasing.

Figure 7 illustrates the overall motion of the lifting subsystem. When the main scissors rotate with respect to the link's angular position, θ_{ms}, the height of the upper platform (UP) changes, as shown in Figure 6. Based on Figure 6, the height changes due to the joints on the platform, where the main scissors (MS) are rotating at the hinge joint on the upper platform (UP) and the base platform (BP). During the platform height change, the end of each main scissor slides on the slider, located at both UP and BP.

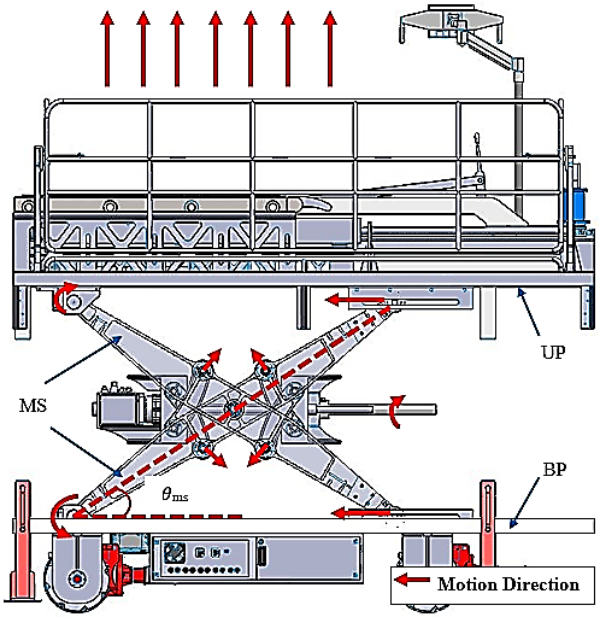


Figure 7. Lifting subsystem working principle.

Figure 8 illustrates that the lifting subsystem consists of similar motions on both the right and left sides of the platform. If the system is configured using a 3D model as depicted in Figure 8, a redundancy issue exists, which may lead to high computational efforts and inaccurate dynamic analysis. To simplify and avoid redundancy, the study utilised the 2D model technique, shown in Figure 9 was used to consider that both sides have similar motions and configurations. Note that all parts were designed based on structural characteristics, and mass properties were assigned before starting the analysis, based on Table 1. Testing revealed no redundancy in the 2D model compared to the 3D model, which had over 10 redundancies.

https://doi.org/10.31436/iiumej.v26i3.3497

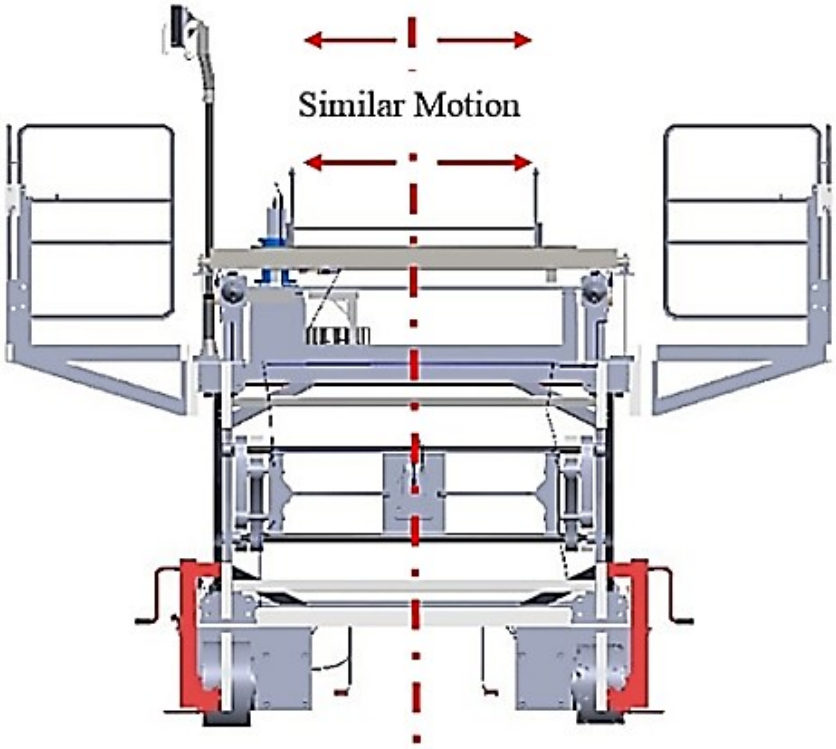


Figure 8. Similar mates and motions.

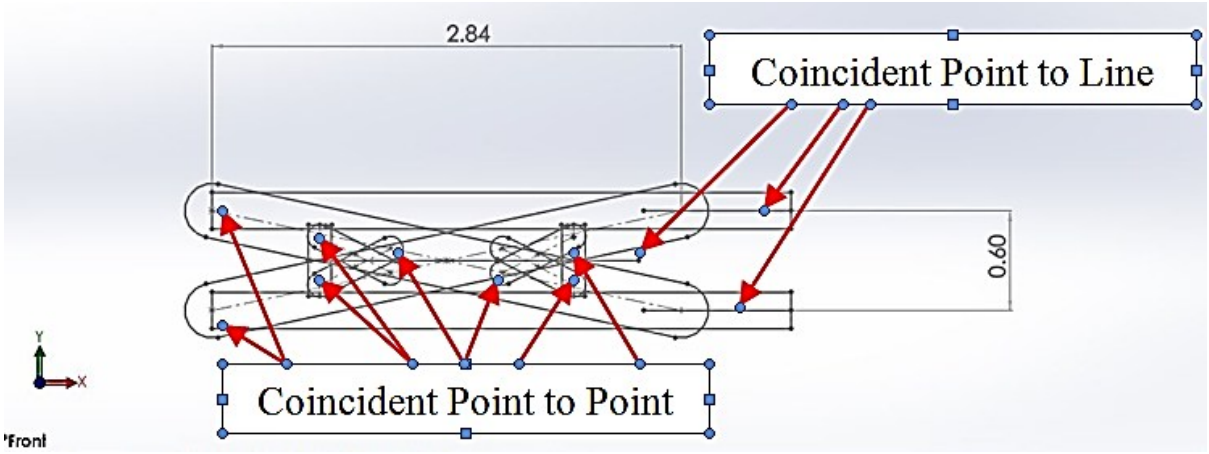
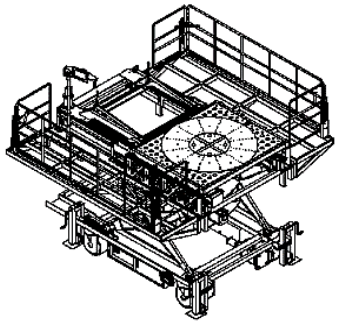
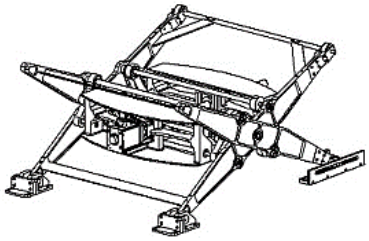
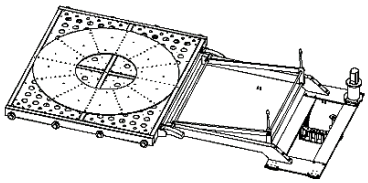
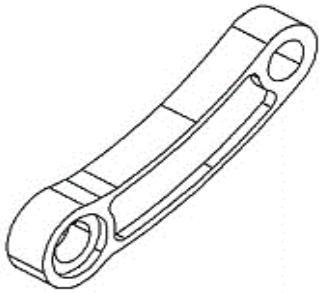
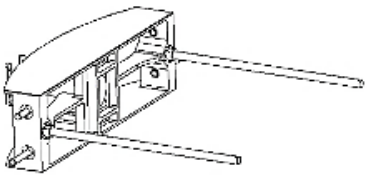
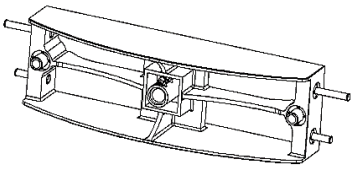


Figure 9. Lifting subsystem 2D model.

Table 1. MAVeP mechanical properties

No.	Structure	Properties
1		Name: All structures Estimated Weight: 7538.5 kg Satellite Weight: 1500 kg Operators' Weight: 200 kg Upper Deck Weight (UP): 1368.5 kg Lower Deck Weight (BP): 2830 kg

https://doi.org/10.31436/iiumej.v26i3.3497

No.	Structure	Properties
2		Name: Lifting Mechanism Estimated Weight: 1840 kg
3		Name: Baseplate Loading Estimated Weight: 1000 kg Baseplate Dimension: 1.94 x 1.64 x 1.00 m
4		Name: Main Scissor Estimated Weight: 92.06 kg Length: 2.9 m Notation: MS
5		Name: Small Scissor Estimated Weight: 37.0 kg Length: 3.85 m Notation: $SS1$, $SS2$, $SS3$ and $SS4$
6		Name: Clamper Estimated Weight: 640 kg Dimension: 1.39 x 0.54 x 0.41 m Notation: $C1$
7		Name: Ball Screw Clamper Estimated Weight: 347 kg Dimension: 1.39 x 0.54 x 0.41 m Notation: $C2$

https://doi.org/10.31436/iiumej.v26i3.3497

2.2. Horizontal motion for baseplate loading subsystem

The baseplate loading subsystem utilises the rack and pinion (RP) mechanism along with the servomotor, as depicted in Figure 10 and 11. Figure 10 provides detailed insight into the working principle of the baseplate loading subsystem. Based on Figure 10, the subsystem is positioned on the upper platform (UP), where the satellite undergoes thermal testing while placed on the baseplate. A clamp ($C3$) secures the baseplate to the mechanism housing the rack and pinion. The baseplate can travel a maximum distance of 3.0 m from its initial position. As shown in Figure 11, this subsystem is straightforward, with the motor ($M2$) directly connected to the pinion, generating rotational motion that is then converted into linear motion of the rack. The combination of the motor and the pinion produces the rotational motion and is converted into the linear motion of the rack.

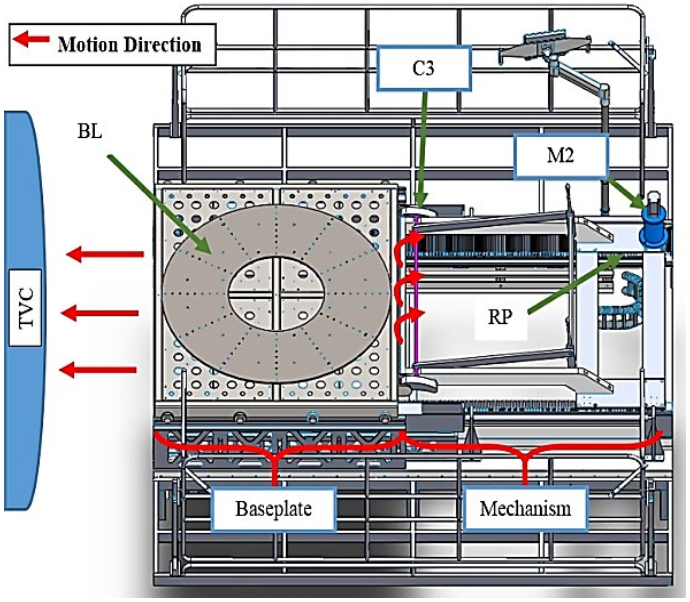


Figure 10. The baseplate loading subsystem configuration (top view).

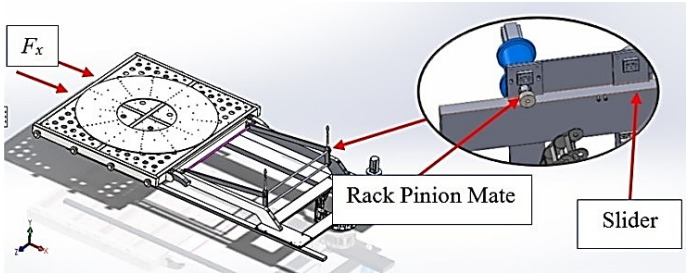


Figure 11. Baseplate loading subsystem 3D model.

Table 2. Selected mechanical attachments

No.	Item	Properties
1	Rack and Pinion	Guide Friction Coefficient: 0.03 Maximum Load: 1280 N Number of Teeth: 20 Pinion Pitch: 0.0637 m
2	Ball Screw	Lead: 0.016 m Maximum Load: 178.9 kN Nut Mass: 17.4 kg Shaft Mass: 40.4 kg

https://doi.org/10.31436/iiumej.v26i3.3497

The 3D model technique was employed for the baseplate loading subsystem due to its simplicity. This method facilitated the implementation of additional mating configurations using mechanical mates in SOLIDWORKS, enabling the system to be set up effectively. Mechanical mates offer various options, including cam, slot, gear, hinge, screw, and rack-and-pinion mates. Hence, the rack and pinion properties listed in Table 2 were incorporated into the motion study. The overall configuration is depicted in Figure 11, and no motion redundancy was found.

3. VELOCITY PROFILE DESIGN

Before analysing the dynamic behaviour of MAVeP, it is essential to elucidate the velocity profile designs utilized for MAVeP operation and dynamic analysis. The researchers detail the speed requirements for each MAVeP subsystem in [1]. For the MAVeP operation, a low velocity is preferred as the operators will be on the platform during the process (loading and unloading). During MAVeP operation, a preference is given to low velocity, as operators will be on the platform during the satellite testing. High velocity can result in sudden movements (jerk), posing safety risks. To mitigate these risks, a velocity profile is employed to regulate acceleration and prevent sudden movements, as detailed in [30, 31], thereby reducing forces exerted around the motor and minimizing motor torque. This velocity profile aims to ensure smooth, fast, and precise motion, avoiding jerking at the start or stop of MAVeP movement, as discussed in [12, 32]. A commonly used approach for designing the velocity profile is the trapezoidal profile, as outlined in [31, 33]. Power consumption can be reduced by segmenting the total operating time into thirds, given that power is inversely proportional to time. Minimizing angular acceleration, a in Eqs. (1)–(3) can reduce the torque and power the servomotor requires. Figure 12 shows a typical velocity profile.

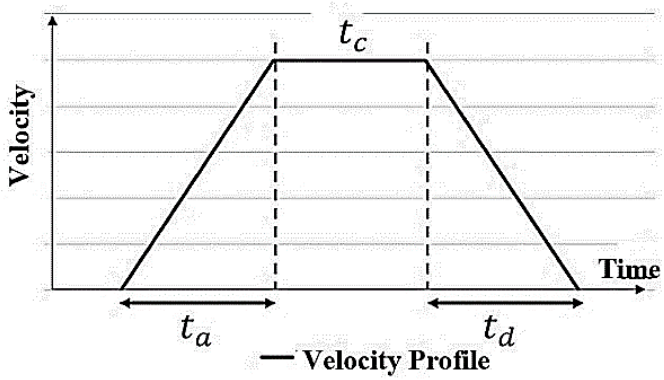


Figure 12. General trapezoidal velocity profile.

First, the motor torque, T_M can be defined as

$$T_M = SF(T_T) \tag{1}$$

$$T_T = T_a + T_L + T_d \tag{2}$$

$$T_a = J_T a \tag{3}$$

where, T_T, T_a, T_L, T_d, J_T and SF denote the total torque, acceleration torque, constant velocity torque, deceleration torque, system moment of inertia, and safety factor, respectively. The general formulae for designing the velocity profile can be described in Eq. (4):

$$D_{Total}(t) = D_a(t) + D_c(t) + D_d(t) \tag{4}$$

https://doi.org/10.31436/iiumej.v26i3.3497

where D_{Total} is the total distance, D_a is the acceleration distance, D_c is the constant velocity distance and D_d is the deceleration distance in the time domain, t. The total distance can be expanded as in Eqs. (5)-(6):

$$D_{Total}(t) = \frac{V_{max}(t)(T_a)}{2} + V_{max}(t)(T - t_a - t_d) + \frac{V_{max}(t)(t_d)}{2} \quad (5)$$

$$D_{Total}(t) = V_{max}(t)(T) - V_{max}(t)(t)\left(\frac{t_a + t_d}{2}\right) \quad (6)$$

where, $V_{max}(t), T, t_a, t_d$ and t_c represent the maximum velocity, total required time, acceleration time, deceleration time, and constant velocity-time, respectively. Then Eq. (6) is rearranged to obtain Eq. (7).

$$V_{max}(t) = \frac{D_{Total}(t)}{T - \frac{t_a + t_d}{2}} \quad (7)$$

t_a and t_d can be defined as shown in Eq. (8). When substituted into Eq. (7), this can be simplified as in Eq. (9), enabling the determination of the total travel time, T in terms of the total travel distance, $D_{Total}(t)$ and maximum velocity, $V_{max}(t)$.

$$\mathrm{t_a = t_d = t_c = \frac{1}{3}T} \quad (8)$$

$$\mathrm{T = \frac{3D_{Total}(t)}{2V_{max}(t)}} \quad (9)$$

The total travel time, T can be computed by substituting the total distance, $D_{Total}(t)$ and maximum velocity, $V_{max}(t)$ in Eq. (9). As T is obtained, the values of t_a, t_d and t_c can be known by solving Eq. (8) to design the velocity profile, as shown in Figure 12. This section has focused on the preliminary design of the velocity profile. The following subsections describe the velocity profile for each subsystem.

3.2. Velocity profile for lifting subsystem

The lifting subsystem has two-speed modes called high and low-speed modes, as stated in [1]. The velocity of the clamper, C1 has been first determined to produce the prescribed speed of the upper platform, UP and the velocity of the clamper is recorded as tabulated in Table 3. To analyse the lifting subsystem's dynamic behaviour, the velocity profiles for the servomotors of the lifting subsystem, as shown in Figure 13, are then developed. These profiles, generated to achieve the prescribed speed of the clamper, C1 in Table 3, are applied to the lifting subsystem in both high and low-speed modes.

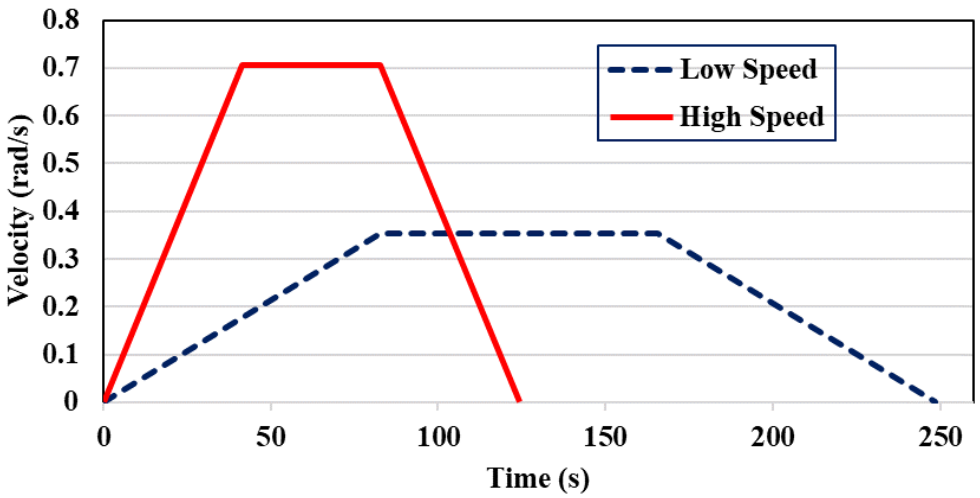


Figure 13. The velocity profiles for the servomotor of the lifting subsystem.

Table 3. Speed requirements for the lifting subsystem

Speed (m/s)	$C1$	UP
High	0.0018	0.0083
Low	0.0009	0.0042

3.2. Velocity profile for baseplate loading subsystem

For the baseplate loading subsystem, a velocity profile was developed explicitly for the servomotor, $M2$, as in Figure 10, to achieve the predetermined speed of the baseplate mechanism (RA), as documented in [1] and detailed in Table 4. The baseplate of MAVeP should move with only one speed, which is 0.0042 m/s. To achieve the speed of the baseplate servomotor, $M2$, should operate at 0.1386 rad/s. Hence, the velocity profile in Figure 14 was generated based on this servomotor speed for the dynamic analysis.

Table 4. Speed assigned for the baseplate loading subsystem

	Servomotor ($M2$)	Baseplate Mechanism (RA)
Speed	0.1386 rad/s	0.0042 m/s

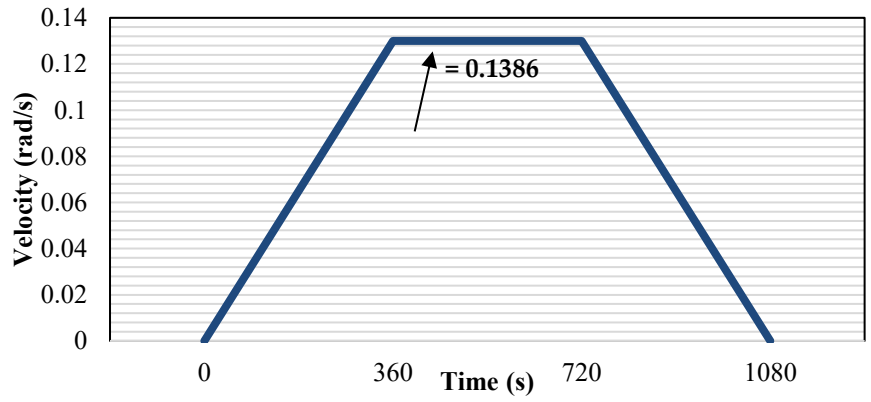


Figure 14. The baseplate loading motor, M2 velocity profile.

4. RESULTS AND DISCUSSION OF DYNAMIC ANALYSIS OF THE SUBSYSTEMS

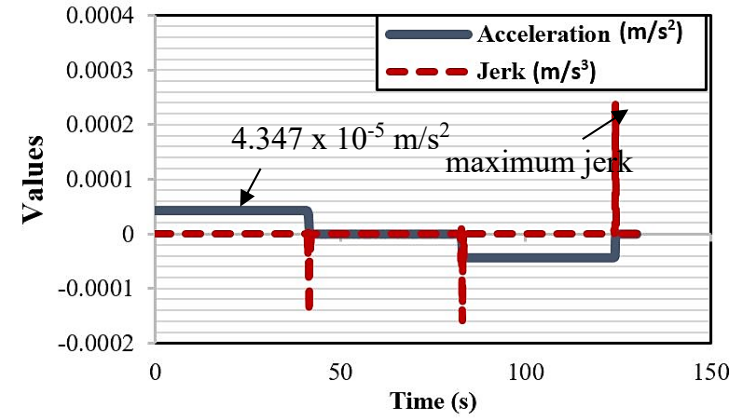
Section 3 comprehensively describes the velocity profile designs for each subsystem in MAVeP. In this section, the dynamic behaviors of each subsystem under the designed velocity profiles shown in Figure 15 and Figure 14, for the lifting (in Figure 9) and baseplate loading (in Figure 11) subsystems, respectively, are analysed. The study was done based on the actual structure specifications, as described in Section 2. This section discusses the dynamic behaviours of the subsystems, such as the acceleration torque, jerk, and velocity.

4.1. Acceleration and jerk analysis

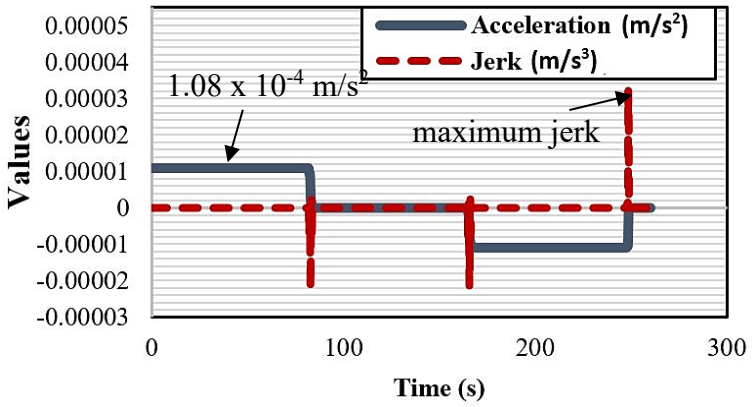
When determining servomotor specifications, the acceleration torque plays a crucial role as it can affect the total torque required by the system. Figure 16 shows the computed acceleration of the designed velocity profiles for the lifting subsystem. Based on the figure, the accelerations are $\pm 4.347 \times 10^{-5}$ m/s^2 for the high speed and $\pm 1.08 \times 10^{-4}$ m/s^2 for the low speed. By referring to Eq. (2), the acceleration torque influences the total torque. However, in this study, the acceleration torque can be disregarded in total torque measurement due to the small acceleration and deceleration of the lifting subsystem (approximately between 10^{-4} m/s^2 and

https://doi.org/10.31436/iiumej.v26i3.3497

10^{-5} m/s^2). Furthermore, the maximum jerk, representing a quick or sudden movement, occurs during acceleration and deceleration at approximately 2.4 × 10^{-4} m/s^3 and 3.1 × 10^{-5} m/s^3 for the high and low speeds, respectively. This jerk rate falls within the low jerk category according to ISO 18738:2003, where the acceptable jerk rate for the elevator ride quality is approximately between 0 and 3.0 m/s^3 [34].



(a) High-speed mode



(b) Low-speed mode

Figure 16. Acceleration and jerk for the lifting subsystem

Figure 17 shows the acceleration and deceleration for the baseplate loading subsystem, both of which are ±1.15 × 10^{-5} m/s^2. The acceleration torque was disregarded because these values are minimal and comparable to those of the lifting subsystem. In this study, the maximum jerk observed was -5 × 10^{-5} m/s^3 at the end of the operation, which falls below the recommended jerk rate in the industry as documented in [34]. The overall findings for both subsystems in terms of acceleration and jerk can be summarised in Table 5. A minimum jerk rate is preferable, as a high jerk rate can lead to platform instability, which may significantly impact the safety of both the operator and any attached equipment, such as satellites. A lower jerk rate also helps reduce mechanical stress and fatigue on the mechanism, extending its operational lifespan.

Table 5. Acceleration and jerk of the subsystems

Subsystems	Lifting	Baseplate Loading
Standard Jerk [34]	0-3.0 m/s^3	0-5.0 m/s^3
High-speed Jerk	2.4 x 10^{-4} m/s^3	5 x 10^{-5} m/s^3
Low-speed Jerk	3.1 x 10^{-5} m/s^3	-
High-speed Acceleration	4.347 x 10^{-5} m/s^2	1.15 x 10^{-5} m/s^2
Low-speed Acceleration	1.08 x 10^{-4} m/s^2	-

https://doi.org/10.31436/iiumej.v26i3.3497

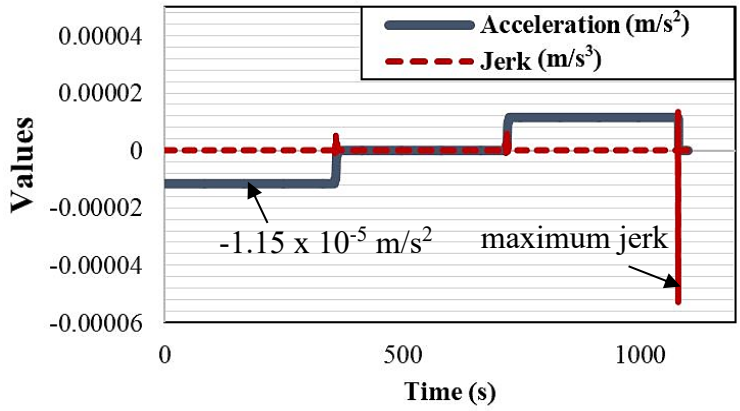
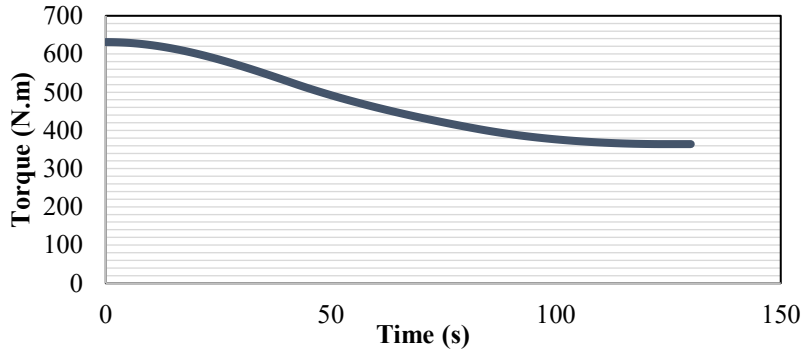


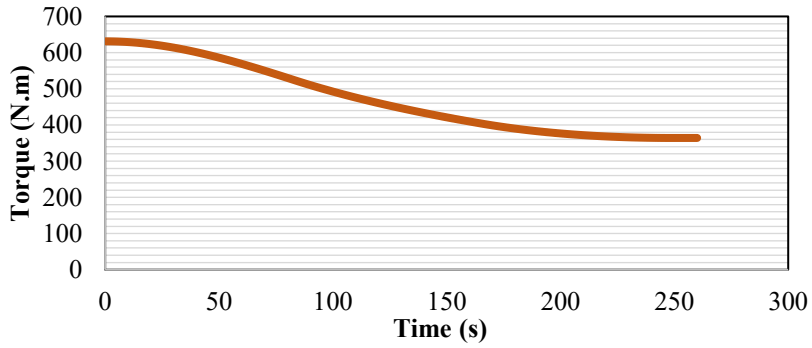
Figure 17. Acceleration and jerk for the baseplate loading subsystem.

4.2. Torque analysis for the subsystems

This study also examined the required torque. It determined the height position of the lifting subsystem and the speed mode (high or low speed) where the highest torque is required. Figure 18 shows the torques required for the servomotor of the lifting subsystem as it transitions from the stowed position to the highest position under the designed velocity profiles. As shown in Figure 18, it can be observed that the maximum torque required for both speeds is similar, which is 631.22 N.m at the highest position and 364.16 N.m at the lowest (stowed) position. The results of this study indicate that speed changes do not affect the maximum and minimum torque required for the servomotor of the lifting subsystem.



(a) High-speed Mode



(b) Low-speed Mode

Figure 18. Servomotor torque requirements for the lifting subsystem

To further validate the results, the findings of this study are compared with those of the study in [24], where the numerical approach was used, as shown in Table 6. From Table 6, the

https://doi.org/10.31436/iiumej.v26i3.3497

result indicates that there is approximately a 28.38% difference. The motor torque was intentionally set higher for safety considerations in the actual system. This ensures that the servomotor does not operate at full power during operation, increasing flexibility and extending its lifespan. Referring to the study in [24], in practical application, the servomotor used for the lifting subsystem of actual MAVeP is 875 N.m, which is only approximately a 27.86% difference compared with the proposed CAD method. Despite a slight difference, the consistency between the proposed method and the real-world application indicates the reliability of our approach. This alignment strengthens confidence in the CAD method's accuracy and its potential for widespread adoption in similar engineering applications.

Table 6. Overview of torque analysis for the lifting subsystem

Speed Mode	**Proposed Method Torque Using 2D Model**	**Calculated Torque Using Numerical Analysis** [24]
Highest Torque (N.m)	631.22	491.69
Lowest Torque (N.m)	346.16	276.78

Figure 19 shows the torque requirements of the servomotor for the baseplate loading system, as determined through dynamic analysis with the proposed 3D CAD model. Based on Figure 19, the maximum torque is approximately 14.68 N.m, where the highest torque occurred during the acceleration, highlighting the significance of torque during this phase. Moreover, the findings show that the proposed velocity profile for the baseplate loading system offers a practical solution to minimize acceleration torque. By implementing this profile, engineers can reduce energy consumption and enhance the system's efficiency. Table 7 summarises the torques required for the baseplate loading subsystem. From the data in Table 7There is approximately only a 9.22% difference compared with the numerical analysis. However, the proposed CAD model provides a more practical and reliable approach compared to the cumbersome and error-prone nature of numerical analysis. This can help practitioners determine the required actuator torque without overspecification, thereby avoiding unnecessary weight and development costs without compromising performance.

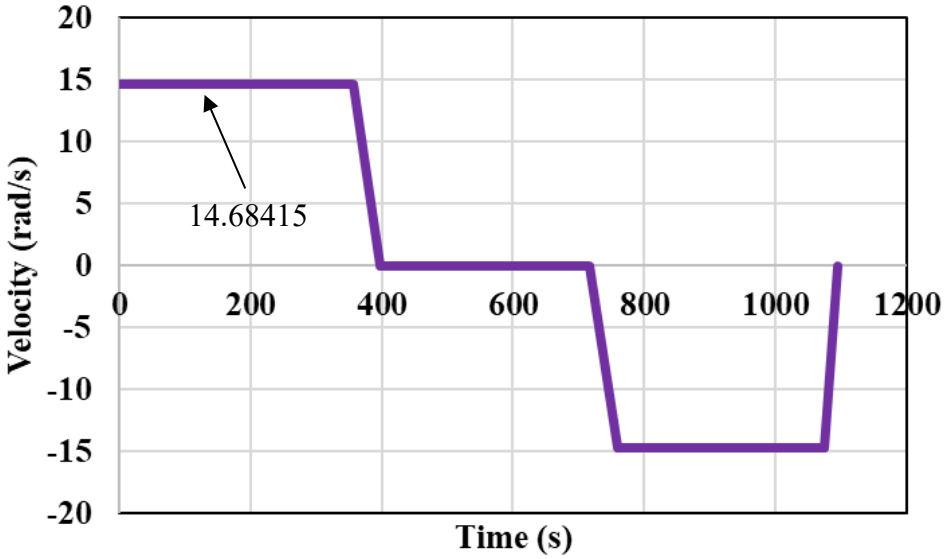


Figure 19. Servomotor torque requirements for the baseplate loading subsystem.

Table 7. Overview of torque analysis for the lifting subsystem

	Proposed Method Torque Using 3D Model	Calculated Torque Using Numerical Analysis
Torque (N.m)	14.68415	18.3793

5. Conclusion

In this study, the dynamic analysis of the lifting and baseplate loading subsystems using the hybrid CAD model was conducted, and the findings were discussed in detail. In summary, this study has shown that the hybrid CAD model, which consists of the 3D and 2D models, has helped practitioners study the dynamic behaviour of the lifting and baseplate loading subsystems of MAVeP without complex numerical formulation. This study also suggests the significance of considering acceleration torque when determining servomotor specifications for systems like the lifting subsystem. Despite the small accelerations and decelerations observed in this study, acceleration torque still impacts the total torque required by the system. Additionally, the study assesses the jerk rate during acceleration and deceleration, which is crucial for evaluating ride quality in systems like aerial work platforms. The observed jerk rates fall within acceptable limits according to ISO standards.

Other than that, the study identified the peak torque demands by determining the height position of the lifting subsystem and the speed mode (high or low speed). Interestingly, this finding suggests that speed changes do not significantly influence the torque demands of the lifting subsystem's servomotor. Such insights are crucial for optimizing servomotor specifications and ensuring efficient performance in practical applications. It is recommended that future research focus on further refining the proposed CAD model and exploring additional optimisation techniques to minimise torque requirements and improve overall system performance.

ACKNOWLEDGEMENT

This project was funded by the Ministry of Science, Technology and Innovation of Malaysia (MOSTI) through the Space Research Agency of Malaysia (ANGKASA) under the SP15-066-0188 grant. The Qatar National Library funded the publication of this article.

REFERENCES

[1] Leng EWL, Salleh N, Salim H, Sabri SF, Ismail M. (2015) Design and development of Motorized Adjustable Vertical Platform (MAVeP) for satellite test facility. In: 2015 Int. Conf. Sp. Sci. Commun. IEEE, pp 424–427.

[2] Norsahperi NMH, Ahmad S, Toha SF, Mutalib MAA. (2022) Design, Simulation and Experiment of PSO-FOPID Controller for Height Position Control of a Scissor. FME Trans, 50(1):46–54.

[3] Norsahperi NMH, Ahmad S, Toha SF, Mahmood IA, Hanif NHHM. (2018) Robustness analysis of fractional order PID for an electrical aerial platform. J Mech Sci Technol, 32(11):5411–5419.

[4] Lu B, Fang Y, Sun N. (2020) Adaptive Output-Feedback Control for Dual Overhead Crane System With Enhanced Anti-Swing Performance. IEEE Trans Control Syst Technol, 28(6):2235–2248.

[5] Chen H, Fang Y, Sun N. (2016) A Swing Constraint Guaranteed MPC Algorithm for Underactuated Overhead Cranes. IEEE/ASME Trans Mechatronics, 21(5):2543–2555.

[6] Spackman H. (1989) Mathematical Analysis of Scissor Lifts.

[7] Corrado A, Polini W, Canale L, Cavaliere C. (2016) To design a belt drive scissor lifting table. Int J Eng Technol, 8(1):515–525.

[8] Oliveira LFA de. (2018) Smart Materials Actuators : Actuated Scissor Lift. University of Porto.

[9] Lee-St.John A, Sidman J. (2013) Combinatorics and the rigidity of CAD systems. Comput Des, 45(2):473–482.

[10] Zhao D-J, Zhao J-S, Yan Z-F. (2016) Planar Deployable Linkage and Its Application in Overconstrained Lift Mechanism. J Mech Robot, 8(2):1–9.

[11] Pan CS, Chiou SS, Kau T-Y, Wimer BM, Ning X, Keane P. (2017) Evaluation of postural sway and impact forces during ingress and egress of scissor lifts at elevations. Appl Ergon, 65:152–162.

[12] Zhang W, Zhang C, Zhao J, Du C. (2015) A Study on the Static Stability of Scissor Lift. Open Mech Eng J, 9(1):954–960.

[13] Singh SK, Deepak SR. (2020) Analytical Reason for Smaller Lateral Sway in Angled-Plane Scissor Linkage. J Mech Robot, 12(5):1–9.

[14] Saxena A. (2016) Deriving a Generalized, Actuator Position-Independent Expression for the Force Output of a Scissor Lift. arXiv Prepr arXiv161110182, :1–18.

[15] Momin GG, Hatti R, Dalvi K, Bargi F, Devare R. (2015) Design, Manufacturing and Analysis of Hydraulic Scissor Lift. Int J Eng Res Gen Sci, 3(2):733–740.

[16] Qu FP, Xue Y. (2014) A High Precision and Convenient Method to Analyze the Mechanics of Scissor Lift Platform under the Action of Combined Loads. Appl Mech Mater, 618:503–508.

[17] Sun LB, Wang RR, Li XX. (2014) Lifting Force Calculation and Safty Analysis of Hydraulic Scissor Lift Platform. Adv Mater Res, 960–961:1450–1454.

[18] Rezaiee-Pajand M, Gharaei-Moghaddam N, Ramezani MR. (2019) A new higher-order strain-based plane element. Sci Iran, 26(4A):2258–2275.

[19] Rezaiee-Pajand M, Gharaei-Moghaddam N, Ramezani M. (2020) Strain-based plane element for fracture mechanics' problems. Theor Appl Fract Mech, 108(March):102569.

[20] Deng M, Ji A, Zhang L, Wang H, Zhao Z. (2022) Lateral vibration behaviors of the straight boom of the aerial work platform in the horizontal linear motion. Autom Constr, 134(December 2021):104095.

[21] Xiao W, Li C, Fan L, Li Q, Tan L. (2023) Dynamic Analysis of the Lifting Arm System in the Integrated Offshore Platform Decommissioning Equipment in Complicated Sea States. Processes. doi: 10.3390/pr11020645.

[22] Pappalardo CM, Del Giudice M, Oliva EB, Stieven L, Naddeo A. (2023) Computer-Aided Design, Multibody Dynamic Modeling, and Motion Control Analysis of a Quadcopter System for Delivery Applications. Machines, 11(4):464.

[23] Shabana AA. (2019) Integration of computer-aided design and analysis: application to multibody vehicle systems. Int J Veh Perform, 5(3):300.

[24] Norsahperi NMH, Ahmad S, Toha SF, Mahmood IA. (2021) Analysis and practical validation on a multi-linkage scissor platforms drive system for the satellite test facilities. Int J Heavy Veh Syst, 28(1):1.

[25] Kyratsis P, Kakoulis K, Markopoulos AP. (2020) Advances in CAD/CAM/CAE technologies. Machines, 8(1):13.

[26] Fan L, Wang J, Xu Z, Yang X. (2022) A Reverse Modeling Method Based on CAD Model Prior and Surface Modeling. Machines, 10(10):905.

[27] Rojas-Sola JI, Río-Cidoncha G Del, Ortíz-Marín R, Trenas-Arbizu J. (2021) An Approach to the Study of a 19th-Century Sugarcane Mill and Steam Engine through CAD Techniques and Mechanical Engineering. Machines, 9(11):295.

[28] Toledano-García AA, Pérez-Cabrera HR, Ortega-Cabrera D, Navarro-Durán D, Pérez-Hernández EM. (2023) Trajectory Generator System for a UR5 Collaborative Robot in 2D and 3D Surfaces. Machines, 11(9):916.

[29] Bhandari B, Manandhar P. (2023) Integrating Computer Vision and CAD for Precise Dimension Extraction and 3D Solid Model Regeneration for Enhanced Quality Assurance. Machines, 11(12):1083.

[30] Huaizhong Li, Le MD, Gong ZM, Lin W. (2009) Motion Profile Design to Reduce Residual Vibration of High-Speed Positioning Stages. IEEE/ASME Trans Mechatronics, 14(2):264–269.

[31] Ha C-W, Rew K-H, Kim K-S. (2013) Robust Zero Placement for Motion Control of Lightly Damped Systems. IEEE Trans Ind Electron, 60(9):3857–3864.

[32] Ha C-W, Lee D. (2018) Analysis of Embedded Prefilters in Motion Profiles. IEEE Trans Ind Electron, 65(2):1481–1489.

[33] Bai Y, Chen X, Sun H, Yang Z. (2018) Time-Optimal Freeform S-Curve Profile Under Positioning Error and Robustness Constraints. IEEE/ASME Trans Mechatronics, 23(4):1993–2003.

[34] (2003) Lift (Elevators)–Measurement of Lift Ride Quality.

https://doi.org/10.31436/iiumej.v26i3.3554

FRONTIER-BASED DETECTION AND SOCIAL FORCE MODEL FOR AUTONOMOUS ENVIRONMENT MAPPING AND NAVIGATION

ASYAM IRSYAD[1], BIMA SENA BAYU DEWANTARA[2*], SETIAWARDHANA[2]

[1]*Electrical Engineering, Politeknik Elektronika Negeri Surabaya, Surabaya, Indonesia*
[2]*Computer Engineering, Politeknik Elektronika Negeri Surabaya, Surabaya, Indonesia*

**Corresponding author: bima@pens.ac.id*

(Received: 29 December 2024; Accepted: 6 August 2025; Published online: 9 September 2025)

ABSTRACT: This paper presents a method for static environment mapping using an autonomous mobile robot by integrating the Social Force Model (SFM) for obstacle avoidance and frontier-based detection for dynamic goal selection. The objective is to enable the robot to autonomously explore unknown environments, avoid obstacles, and generate accurate maps. The system was developed and tested in the Gazebo simulator, with RViz employed to visualize real-time sensor data, including lidar and odometry. Two frontier point selection strategies were evaluated: one based solely on the nearest distance, and another incorporating both distance and orientation. Experimental results demonstrate that the first strategy enabled the robot to complete the mapping task in an average of 10.66 minutes over a distance of 76.37 meters. In contrast, the second strategy, although more directionally aware, resulted in a longer duration of 14 minutes and a travel distance of 104.08 meters. These outcomes highlight the impact of map update delays and suboptimal frontier point selection on overall navigation efficiency. The findings suggest that while the system exhibits promising autonomous navigation capabilities, further improvements in frontier map update speed and goal point optimization are necessary to enhance path efficiency and reduce unnecessary movement.

ABSTRAK: Kertas ini membentangkan kaedah pemetaan persekitaran statik menggunakan robot mudah alih autonomi dengan mengintegrasikan Model Daya Sosial (SFM) bagi mengelak halangan dan pengesanan berasaskan sempadan bagi pemilihan matlamat dinamik. Objektifnya adalah untuk robot meneroka persekitaran yang tidak diketahui secara autonomi, mengelak halangan, dan menjana peta dengan tepat. Sistem ini dibangunkan dan diuji dalam simulator Gazebo, manakala RViz digunakan bagi memvisualisasikan data sensor masa nyata termasuk input lidar dan odometri. Dua strategi pemilihan titik frontier telah dinilai: satu berdasarkan jarak terdekat sahaja, dan satu lagi yang mengambil kira jarak serta orientasi. Dapatan eksperimen menunjukkan bahawa strategi pertama membolehkan robot melengkapkan tugas pemetaan dalam purata masa 10.66 minit dengan jarak perjalanan 76.37 meter. Manakala, strategi kedua, walaupun lebih peka terhadap arah, memerlukan masa yang lebih lama iaitu 14 minit dengan jarak perjalanan 104.08 meter. Dapatan kajian ini menunjukkan bahawa kelewatan dalam mengemas kini peta dan pemilihan titik frontier yang kurang optimum memberi kesan kepada kecekapan navigasi keseluruhan. Walaupun sistem ini menunjukkan kebolehan navigasi autonomi yang baik, penambahbaikan dalam kelajuan kemas kini peta frontier dan pengoptimuman pemilihan titik matlamat diperlukan bagi meningkatkan kecekapan laluan dan mengurangkan pergerakan yang tidak perlu.

KEYWORDS*: Autonomous Navigation, Social Force Model (SFM), Simultaneous Localization and Mapping (SLAM), Frontier Detection, Robot Exploration.*

https://doi.org/10.31436/iiumej.v26i3.3554

1. INTRODUCTION

Robotic technology has greatly sped up its development in response to the growing need for more difficult tasks and roles [1-6]. To fulfill the need for human-robot interaction, psychological and social factors are increasingly being considered [7-10]. Along with hardware systems that function well, software and algorithms, such as those used in detection and analysis, navigation, object avoidance, localization, and mapping, are also crucial.

One essential capability that a mobile robot must possess is the ability to navigate within a specific environment. Typically, the environment where the robot will operate can be mapped beforehand. When applied in environments with static obstacles (such as walls and furniture) and dynamic ones (including humans, animals, and other moving objects), a human-operated robot must first map the area [11, 12]. Making maps can also be done by a mobile robot on its own in places where there are only static barriers [13]. Consequently, the majority of robots encounter difficulties navigating unfamiliar environments [14, 25-27].

Furthermore, mapping the area is not difficult for either the robot or the operators in structured and safe environments. However, in unstructured or dangerous areas, a human user might not be present, so the robot needs to be able to map the area on its own. Additionally, the need for a workforce is also something to consider [15]. This study combines Simultaneous Localization and Mapping (SLAM) technology with local navigation tools, particularly the Social Force Model (SFM) [9, 17, 21-23]. This enables a mobile robot to navigate autonomously while identifying its location and interpreting its surroundings. Being able to move independently means that the robot can avoid both static and moving objects while it is working.

SLAM-toolbox was selected in this research due to its robust features and practical advantages [20]. SLAM-toolbox not only simplifies the implementation of SLAM by providing a ready-to-use framework, eliminating the need to develop SLAM algorithms from scratch, which can be highly time-consuming, but it also incorporates occupancy grid mapping. This feature is particularly beneficial as it allows seamless integration with additional algorithms, such as frontier-based exploration, to enhance mapping and navigation capabilities in diverse environments [14, 25-27]. The reason SFM was chosen is that it is easy to use, can accurately model social interactions, can avoid collisions, works well in dynamic environments, and can adapt to various situations [10].

The main contributions of integrating the Social Force Model (SFM) navigation method with SLAM-toolbox and frontier-based exploration are:

- The integration of SFM and SLAM-toolbox enables the robot to autonomously navigate dynamic environments, such as areas with human activity or moving objects, by accounting for social interactions. This ensures that the robot can not only build and maintain an accurate map of the environment but also safely avoid collisions with humans and dynamic obstacles.
- SLAM-toolbox provides a robust framework for real-time map generation and updates using occupancy grids. At the same time, the inclusion of frontier-based exploration allows the robot to identify unexplored areas efficiently. Coupled with SFM, the robot adapts to dynamic environmental changes, maintaining safety and mapping precision during autonomous navigation and exploration.
- The SFM allows the robot to mimic human-like navigation behavior, maintaining safe distances and moving naturally in crowded or socially sensitive environments. This ensures

https://doi.org/10.31436/iiumej.v26i3.3554

comfortable interactions and reduces potential disruptions in human-robot interaction scenarios.

- The combination of SFM, SLAM-toolbox, and Frontier-based Detection provides the robot with adaptability in diverse scenarios, whether in static or dynamic environments, and enhances its capability to explore both confined and open spaces with varying population densities.
- This integration significantly increases the robot's autonomy by enabling it to navigate, explore, and interact with dynamic and complex environments without human intervention. The robot can dynamically adjust its behavior while simultaneously updating its map and planning exploration paths to maximize coverage.

The remainder of this paper is organized as follows. Section 2 describes the mobile robot navigation, frontier detection, localization, and mapping system simultaneously. Section 3 delivers the experimental results and discussion. Finally, the conclusion and future work are presented in Section 4.

2. LITERATURE REVIEW

2.1. Social Force Model (SFM)

The Social Force Model (SFM) predicts human movement in dynamic environments [17]. It considers "social forces," which reflect not just physical forces but also the desire to achieve goals while avoiding obstacles and interacting with others. This model is particularly effective for socially aware navigation in robots.

In robotics, the SFM helps robots avoid fixed and moving obstacles by calculating the necessary force to navigate. The total navigation force (F_{nav}) refer to Eq. (1) is the sum of two components: the force to move toward the goal (F_g) refer to Eq. (2) and the force from static obstacles (F_s). Here, v^0 represents the desired velocity of the robot, while v is the current velocity. The sum of physical and social repulsion is a static force, referred to as Eq. (3). This force keeps the robot away from a stationary object.

$$F_{nav} = F_g + F_s \tag{1}$$

$$F_g = m \cdot a = m\frac{v^0 - v}{t} \tag{2}$$

$$F_s = f_{soc}^s + f_{phy}^s \tag{3}$$

$$f_{soc}^s = k^s exp + \left(\frac{r_R - d_R^s}{\psi^s}\right) \cdot \overrightarrow{e_s} \tag{4}$$

$$f_{phy}^s = k^s \cdot (r_R - d_R^s) \cdot \overrightarrow{e_s} \tag{5}$$

The social robot's repulsive force against static obstacles is denoted as f_{soc}^s refer to Eq. (4), while its physical repulsion force is f_{phy}^s as seen as Eq. (5). The robot's radius of the proxemics area is r_R, its distance to the nearest static obstacle is d_R^s, and a gain factor k^s determines future feedback. The effective distance value for avoiding dynamic obstacles is ψ^s, and $\overrightarrow{e_s}$ represents the vector direction of the working force. The vector of SFM is represented in Fig. 1.

https://doi.org/10.31436/iiumej.v26i3.3554

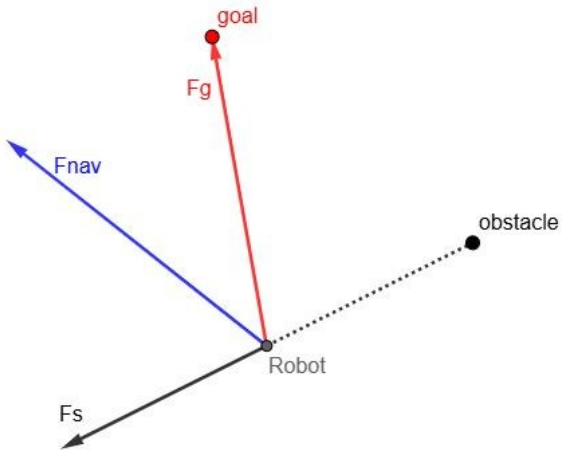


Figure 1. The forces that operate within the SFM framework.

2.2. Simultaneous Localization and Mapping

Simultaneous Localization and Mapping, or SLAM, is a way for robots or self-driving cars to make maps and figure out where they are on the map simultaneously. In this case, SLAM uses algorithms, mathematical models, and sensor data to assist robots in determining their position and actions within unfamiliar environments. Besides that, the SLAM software system can also figure out where the robot is in its surroundings [18].

SLAM-toolbox is a widely used framework in robotics for simultaneous localization and mapping (SLAM), enabling a robot to construct a map of its surroundings while simultaneously determining its position within the environment. SLAM-toolbox is particularly effective for applications requiring high-precision 2D or 3D mapping, especially in structured indoor settings with static features [20]. The key processes and formulations implemented in SLAM-toolbox are detailed in Algorithm 2.

2.3. Robot Model

The Differential Drive Mobile Robot (DDMR) Kinematic function determines the wheel speed of the robot, which requires kinematics, allowing it to calculate the wheel speed difference between the left and right wheels using the robot's coordinate distance to the goal. In the DDMR robot kinematic system with 2 wheels and 1 free-wheel front wheel, to calculate the speed of the robot, the formula for linear speed and angular speed is as Eq. (6 - 9) follows:

$$V_R = \omega_R . r_R \tag{6}$$

$$V_L = \omega_L . r_L \tag{7}$$

$$v = \frac{V_R + V_L}{2} \tag{8}$$

$$\omega = \frac{\omega_R - \omega_L}{2} \tag{9}$$

V_R and V_L show how fast the right and left wheels are moving in a straight line, and R and L show how fast the wheels are turning in a circle. Using the coordinate points (x, y) of the end point to find the wheel speeds, the inverse kinematics of each wheel's speed is used. Here, the input θ, which tells the robot which way to go, will be set separately. The values of x, y and θ are obtained as in Eq. (10 – 12).

$$\dot{x} = v . \cos(\theta) \tag{10}$$

https://doi.org/10.31436/iiumej.v26i3.3554

$$\dot{y} = v.\sin(\theta) \tag{11}$$

$$\dot{\theta} = \omega \tag{12}$$

An illustration of equations on the DDMR Robot can be seen in Fig. 2.

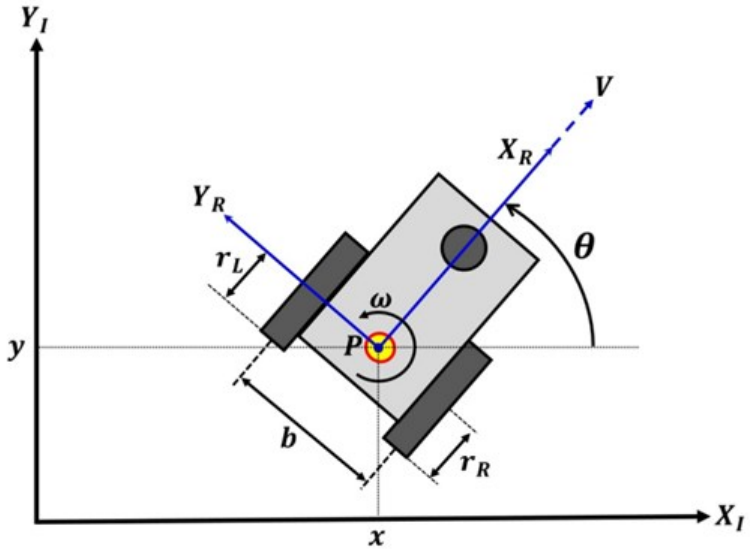


Figure 2. Illustration of Equations on the DDMR Robot [24].

2.4. Frontier Detection

Frontier detection with an occupancy grid is a method for pinpointing cells at the interface between mapped and unmapped regions inside a grid. Frontier cells are crucial for devising exploration routes, as they delineate the boundary between known (blocked or free) and unknown regions [14, 25-27].

Algorithm 1: isFrontierCell

Input: Matrix B, coordinates (x, y)
Output: True if (x, y) is a frontier cell, otherwise False

```
if (x, y) is out of bounds or B[y,x] ≠ −1 then
    return False;
Define neighbors as the 8 surrounding cells (including diagonals);
hasZeroNeighbor ← False;
allNeighborsUnknown ← True;
for each (nx, ny) in neighbors do
    if (nx, ny) is within bounds then
        neighborValue ← B[ny, nx];
        if neighborValue == 0 then
            hasZeroNeighbor ← True;
        if neighborValue ≠ −1 then
            allNeighborsUnknown ← False;
return (hasZeroNeighbor AND NOT allNeighborsUnknown);
```

https://doi.org/10.31436/iiumej.v26i3.3554

Algorithm 2: Detect Frontiers

Input: A matrix B of size $width \times height$
Output: List of centroids of strongly connected components

```
Create empty set frontier;
for x ← 0 to B.width − 1 do
    for y ← 0 to B.height − 1 do
        if isFrontierCell(B, x, y) then
            Add (x, y) to frontier;
SCC_list ← Kosaraju_DFS(frontier);
centroids ← [];
for each component in SCC_list do
    if component.size > 8 then
        centroid ← calculateCentroid(component);
        Add centroid to centroids;
return centroids;
```

The algorithm detects frontier cells in a matrix, which are cells with a value of -1 that border empty areas (value 0) and have at least one known neighbor (not -1), as shown in Algorithm 1. It then identifies strongly connected components (SCCs) using Kosaraju's Depth-First Search and computes the centroid of each component with more than 8 cells (Algorithm 2). These centroids represent candidate destinations, and the final goal is selected based on the chosen destination selection strategy.

2.5. Related Works

Several previous studies have explored the use of the Social Force Model (SFM) in combination with various supporting technologies to improve mobile robot navigation. Table 1 summarizes representative research efforts that utilize SFM, highlighting their main objectives and key findings. These studies demonstrate the adaptability and effectiveness of SFM across diverse environments and applications.

Table 1. Summary of related works using SFM for mobile robot navigation

Authors (Year)	Supporting Technologies	Primary Objective	Key Findings / Results
Rifqi et al. (2021)	Fuzzy Inference System (FIS), LiDAR	Adaptive navigation for healthcare robots in social environments	79.9% navigation success; safe behavior via adaptive SFM parameters.
Dewantara & Miura (2016)	Reinforcement Learning (Q-learning	Generate socially aware guiding behaviors for a guide robot	Reduced over-reactive behavior; achieved smoother, safer, and more comfortable guidance.
Dewantara & Ariyadi (2021)	Fuzzy Inference System (FIS)	Adaptive navigation control for robot soccer in dynamic environments	Improved reactivity and responsiveness; adaptive SFM parameters outperformed constant ones.
Ramadhan et al. (2023)	Fuzzy Inference System (FIS), Genetic Algorithm (GA)	Optimize SFM parameters adaptively for mobile robot navigation	Smoother robot travel (reduced heading deviation); reached goal 1.6 sec faster than fixed parameters.

https://doi.org/10.31436/iiumej.v26i3.3554

Inspired by the adaptive use of SFM in the aforementioned studies, this paper proposes a system that combines SFM with frontier-based exploration and SLAM for enhanced autonomous mapping and navigation.

3. METHODOLOGY

3.1. Proposed System

The proposed system design is illustrated in Fig. 3. The lidar sensor is used to measure object distances within the environment. The lidar data is processed to determine distances to static objects. The information from static objects is then fed into the Social Force Model (SFM) module to compute the robot's heading and speed, ensuring safe and human-aware navigation. Simultaneously, the static object data is processed by the SLAM-toolbox module, which localizes the robot and generates an occupancy grid map of the environment. The Frontier-based Detection module further utilizes the resulting map to identify unexplored regions, enabling efficient and autonomous exploration of the environment.

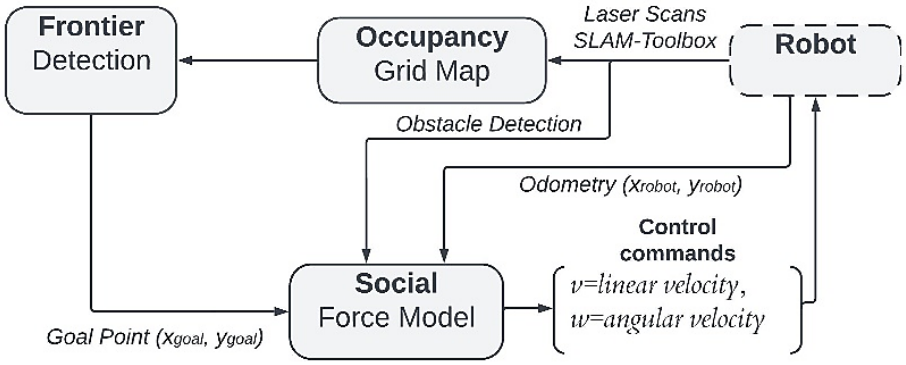


Figure 3. Proposed system

3.2. Simulation

This study is mainly about robot navigation, localization, and mapping. It uses crucial algorithms like the Social Force Model (SFM) for planning local paths, Simultaneous Localization and Mapping (SLAM) for mapping environments, and Frontier Detection for finding goal points. The primary simulation tool for the robot's behavior and environment is Gazebo. RViz is used as the graphical user interface to see how the robot interacts with its environment. In this setting, RViz processes and monitors real-time sensor data, including lidar inputs and odometry readings. This lets people see where the robot is ($R(x, y)$), which way it is facing, and how good the maps made by the SLAM method are. The study concentrates on three primary objectives:

- Navigation: The Social Force Model (SFM) tests the robot's autonomous movement. The local planner tool SFM helps the robot avoid impediments on its way to its destination. Social influences regulate the robot's speed and path to the goal to mimic human movement.
- SLAM/Mapping: This goal combines localization and mapping so the robot can find itself and map its surroundings. Independent navigation requires SLAM to provide the robot with a worldview to plan its subsequent actions.

Frontier detection finds and investigates unexplored environmental areas. Frontier Detection finds known and undiscovered map edges. Recognizing these edges enables the robot to determine where to investigate next, thereby expanding the mapped areas. This function

https://doi.org/10.31436/iiumej.v26i3.3554

allows the robot to explore new places and avoid previously visited ones, thereby making the map more comprehensive.

The combination of Gazebo and RViz provides a controlled and flexible platform to simulate ideal sensor data in a realistic virtual environment. This approach enables the validation of navigation and SLAM algorithms without the risks and limitations of physical hardware, supporting efficient research in customizable simulation scenarios.

3.3. Experimental Setup

All experiments were conducted on a system with the following specifications: Intel NUC, Core i5-1135G7 (4.2 GHz), 8 GB DDR4 RAM, Iris Xe Integrated Graphics, Ubuntu 22.04 LTS, and a 128 GB NVMe M.2 SSD. The development environment included Visual Studio Code (VSCode), OpenCV 4.9.0 libraries, as well as the Gazebo and RViz 3D simulation tools. Gazebo and RViz were integrated with ROS2 for real-time robotic control and visualization. For experimental purposes, the specified SFM parameters are presented in Table 2.

Table 2. SFM Parameters

Parameters	Value
Gain	3
Robot mass	1.5 kg
Max. speed	0.35 m/sec
Proxemic type	half circle
Proxemic radius	1.3 m
Effective range	1 m

3.4. Experimental Scenario

In this research, using the TurtleBot3 Waffle model, as shown in Fig. 4, which is provided by the community in the ROS 2 framework [28]. This model was chosen because it is well-integrated into the ROS 2 ecosystem and comes equipped with the 360° lidar sensor required for this research. Using the available model can avoid the time-consuming process of designing the robot from scratch, allowing for a greater focus on algorithm development and testing in simulations. In addition, TurtleBot3 is classified as a Differential Drive Mobile Robot (DDMR), which utilizes two independent wheels to drive the robot, providing precise control over its movement and making it an ideal platform for research and development in robotics.

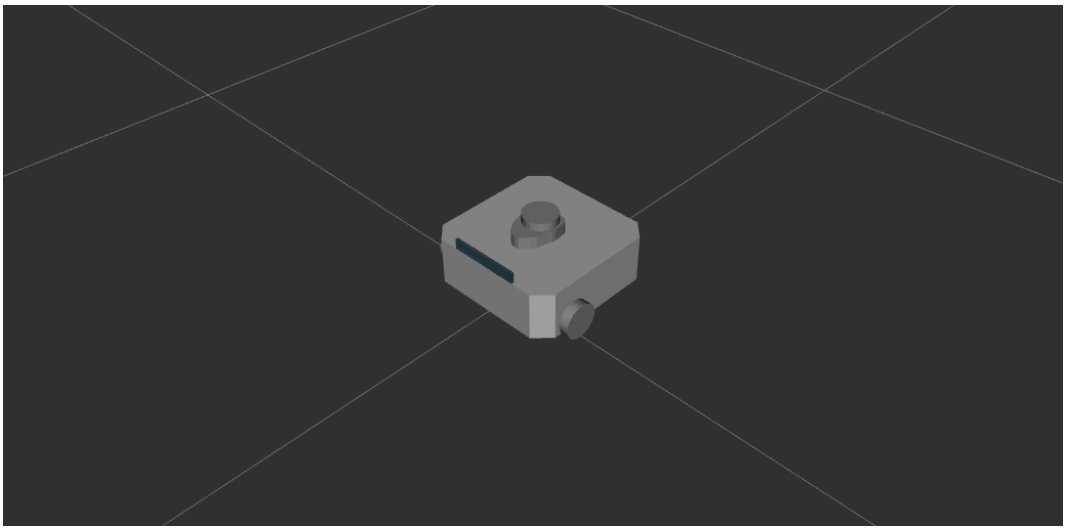


Figure 4. Turtlebot3 model used in the simulation.

https://doi.org/10.31436/iiumej.v26i3.3554

Fig. 5 shows the experimental scenario where the robot is placed in a room, as seen in the image, with an area of approximately 155 m^2. The robot's objective is to map the entire room without colliding with obstacles, then measure the distance traveled to complete the mapping and the time required to perform the mapping.

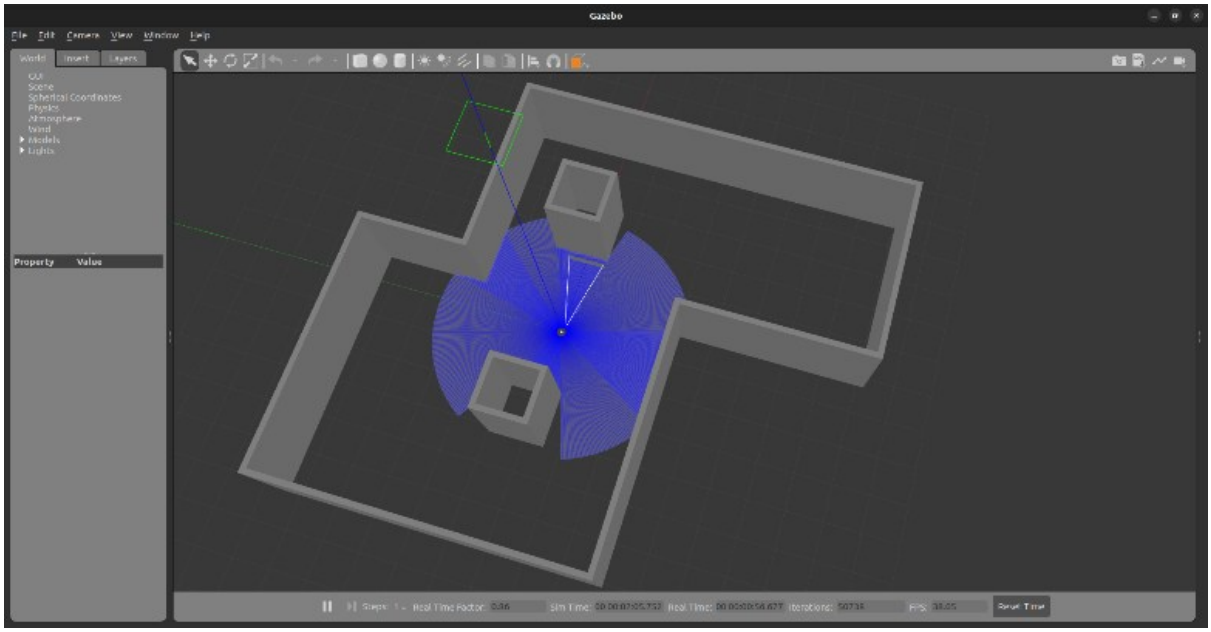


Figure 5. Experiment layout room scenario.

4. EXPERIMENTAL RESULTS AND DISCUSSION

Frontier detection produces candidate frontier points that represent potential goal locations for the robot. In the scenario illustrated in Fig. 6, the selection strategy relies solely on identifying the nearest point from the robot's current location. While this approach is computationally simple, it often results in inefficient trajectories, as the robot consistently chooses the closest frontier point without considering direction or overall path efficiency. This can cause the robot to perform frequent orientation changes, including abrupt turns of up to 180 degrees. After reaching a goal, the map is updated and new frontier points are generated, prompting the robot to re-evaluate and select the next closest target. Due to this greedy behavior, the resulting path is sometimes unnecessarily convoluted.

The outcomes across three trials under this strategy demonstrate variations in performance. The robot traveled 77.53 m, 61.33 m, and 90.26 m, with corresponding mapping times of 13, 9, and 10 minutes, respectively. On average, the travel distance was 76.37 m, and the mapping time was 10 minutes and 40 seconds.

https://doi.org/10.31436/iiumej.v26i3.3554

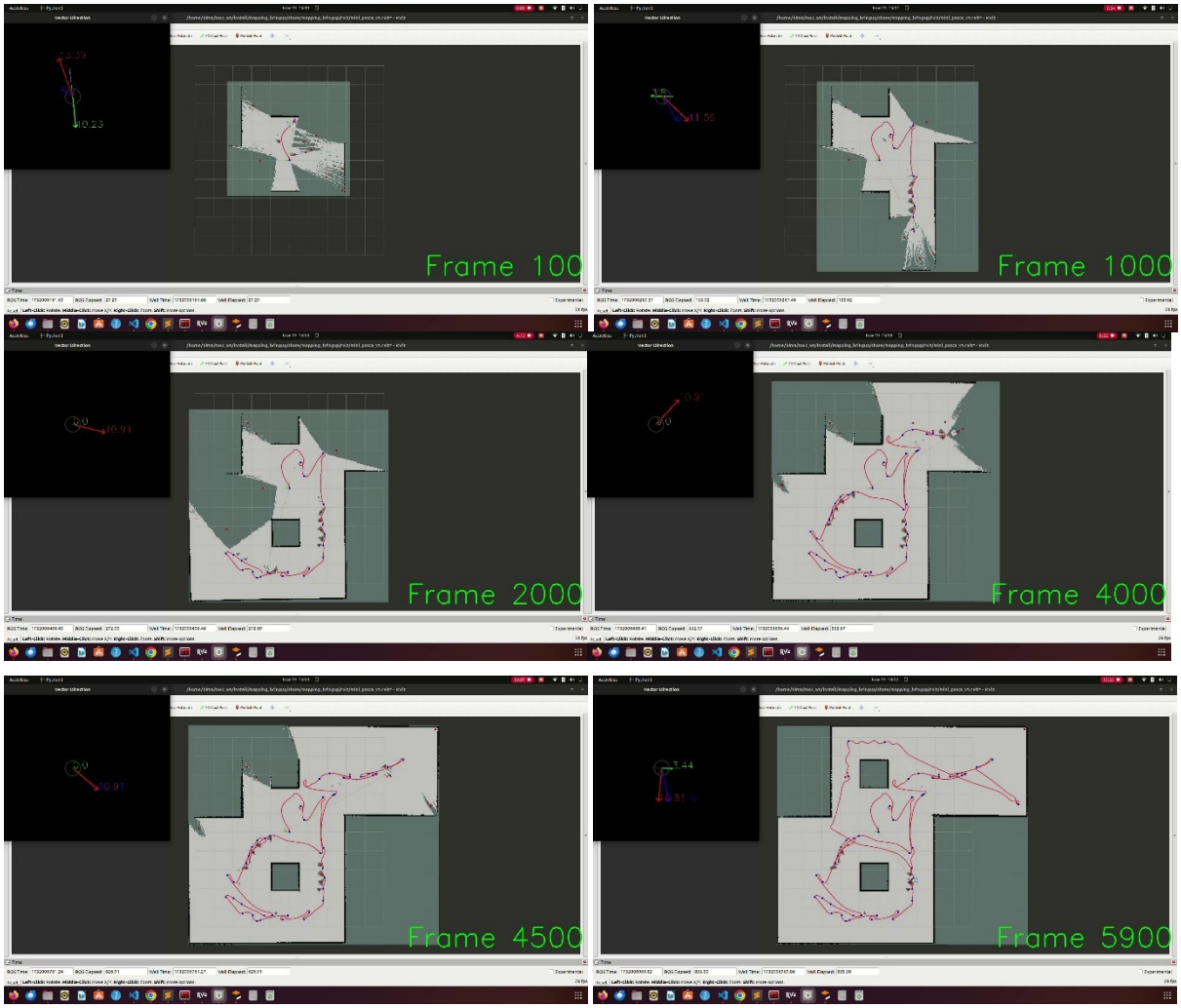


Figure 6. Mapping a room using the nearest distance of frontier points.

In contrast, Fig. 7 shows the result of using a more refined selection strategy that considers both the nearest distance and the closest orientation relative to the robot's current heading. This method aims to reduce unnecessary rotation and improve navigational smoothness. However, in practice, this strategy still suffers from inefficiencies due to timing mismatches between goal arrival and frontier detection updates. Specifically, when the robot reaches a goal point, the system may not yet finish updating the frontier points based on the new map, as the algorithm performs a complete scan of all cells. Consequently, the robot may select a suboptimal goal, one that appears optimal based on outdated map information resulting in a longer and more winding path.

In the three experiments conducted under this method, the robot traveled 99.47 m, 88.75 m, and 124.02 m, with mapping times of 17, 13, and 12 minutes, respectively. The average travel distance was 104.08 m, and the average mapping time was 14 minutes.

https://doi.org/10.31436/iiumej.v26i3.3554

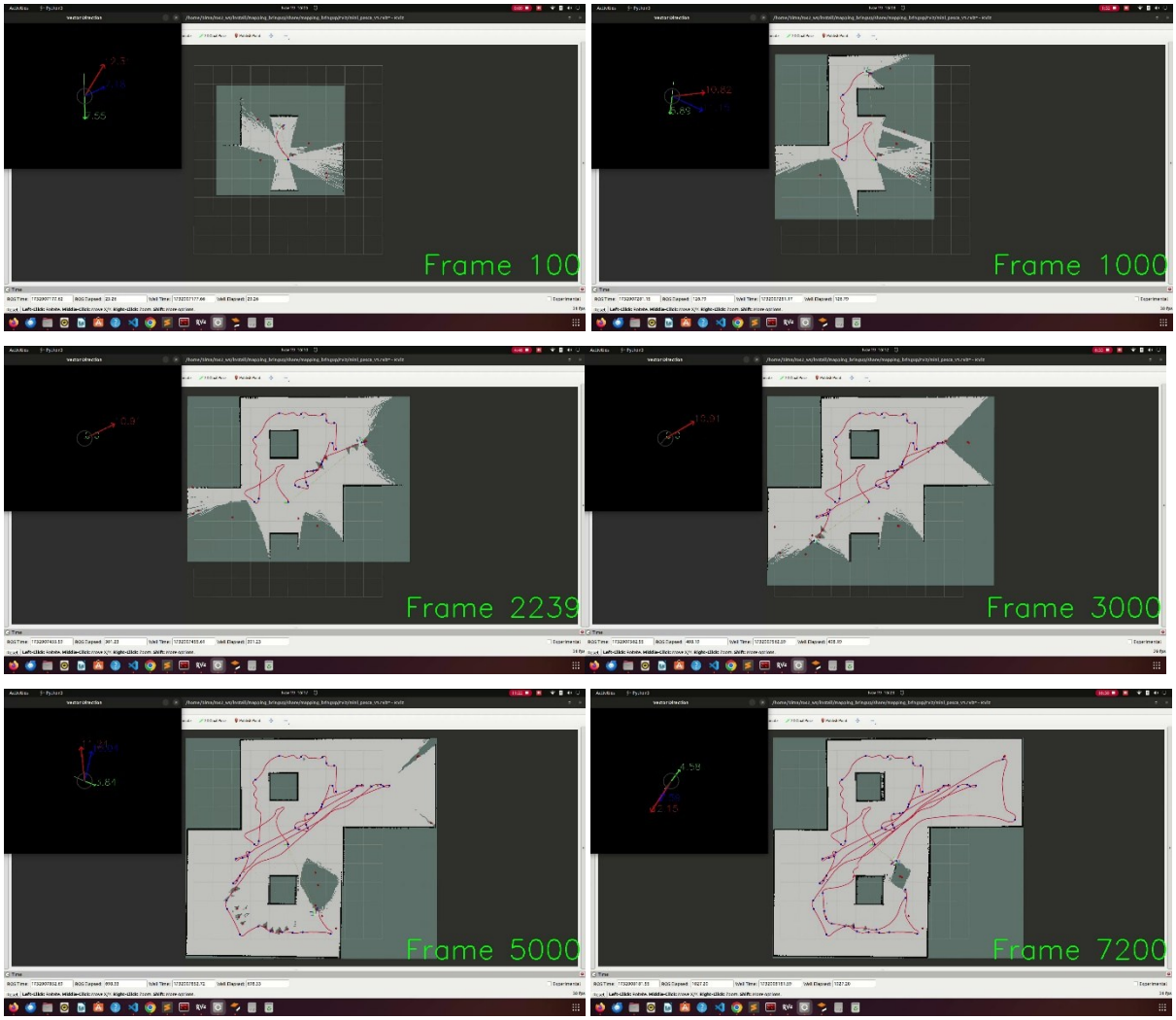


Figure 7. Mapping a room using the nearest distance and nearest orientation of frontier points.

A comparison of the two frontier point selection strategies is summarized in Table 3. The first strategy selects the frontier point based only on the shortest distance to the robot, while the second strategy additionally considers the orientation of the robot toward the point. Each approach was tested in three different experimental runs.

Table 3. Comparison of travel distance and mapping time for different point selection strategies.

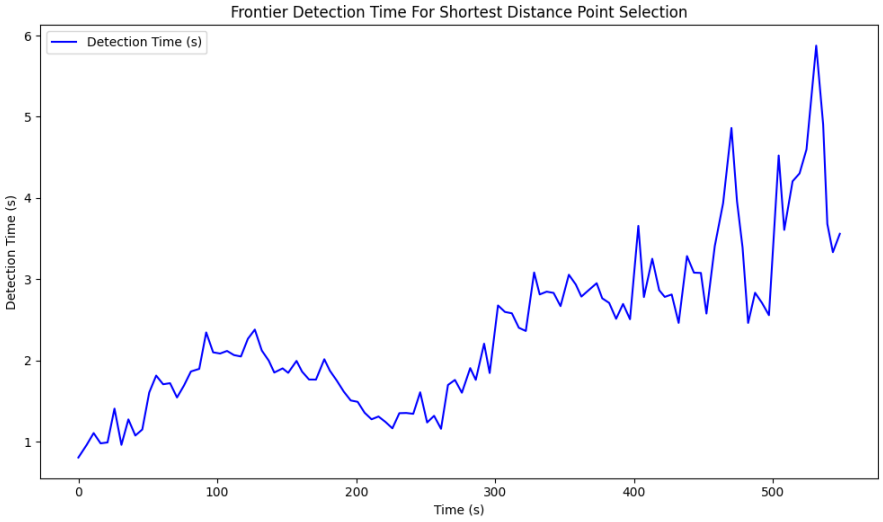
Point Selection	**Experiment (n)**	**Travel Distance (m)**	**Mapping Time (minutes)**
Shortest distance	1	77.53	13
	2	61.33	9
	3	90.26	10
	Avg	**76.37**	**10.66**
Shortest distance & closest orientation	1	99.47	17
	2	88.75	13
	3	124.02	12
	Avg	**104.08**	**14**

https://doi.org/10.31436/iiumej.v26i3.3554

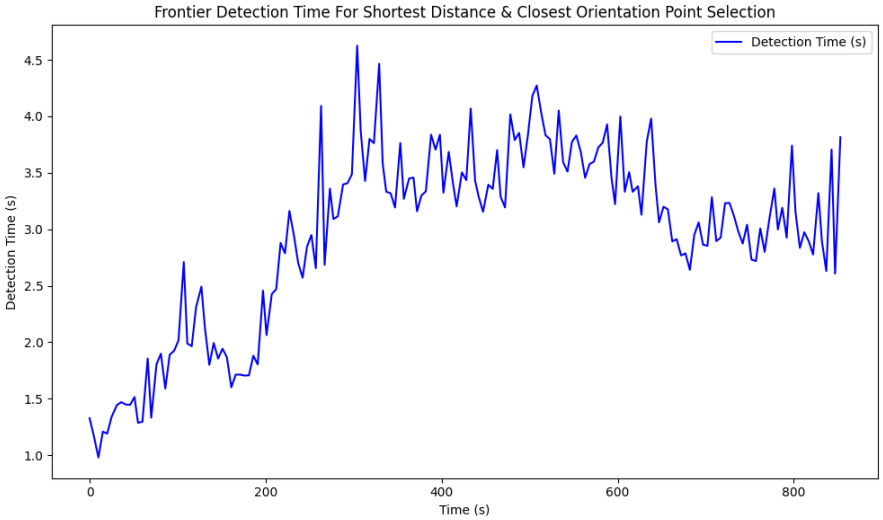
From the table above, it is evident that the strategy considering both distance and orientation tends to result in a longer travel distance and mapping time. Although this approach aims to align the robot's direction with the target, the delayed frontier update occasionally causes the robot to select less optimal points. In contrast, the shortest distance only strategy yields more efficient performance in terms of time and distance, albeit with less smooth trajectories.

The computation time required for frontier detection is also critical. Fig. 8 illustrates the detection time performance throughout the exploration. In the shortest distance scenario (Fig. 8a), the average frontier detection time is 2.34 seconds, with a maximum of 5.87 seconds and a minimum of 0.80 seconds. Meanwhile, in the shortest distance and closest orientation scenario (Fig. 8b), the average detection time increases to 2.95 seconds, with a maximum of 4.62 seconds and a minimum of 0.97 seconds.

These results indicate that as the map grows and the number of cells increases, the time needed to detect frontiers becomes longer, especially when the algorithm scans all grid cells (as described in Algorithm 1). Consequently, the robot may select a suboptimal or outdated frontier point before the detection algorithm completes processing the updated map.



(a) Shortest distance point selection scenario.



(b) Shortest distance and closest orientation point selection scenario.

Figure 8. Detection time for the frontier detection algorithm over time during exploration.

5. CONCLUSION

This study integrated the Social Force Model, SLAM Toolbox, and frontier-based detection to enable autonomous map exploration by a differential drive mobile robot. The system was evaluated in a controlled environment using the Gazebo simulator. Two frontier point selection strategies were tested: (1) based solely on the nearest distance, and (2) based on the nearest distance with orientation consideration. Experimental results show that the first strategy enabled the robot to complete the mapping task in an average of 10.66 minutes over a distance of 76.37 meters, while the second strategy required an average of 14 minutes and a distance of 104.08 meters. Although the second strategy incorporated orientation alignment, it resulted in a longer and more convoluted path. This was primarily due to two factors: limited frontier point selection logic and delay in frontier map updates, which occasionally caused the robot to choose outdated or inefficient target points. These findings indicate the need for improvements in both the frontier detection update speed and the frontier point selection strategy. Enhancing the update frequency of the frontier map would allow the robot to make more informed and timely decisions. Furthermore, optimizing the point selection mechanism by balancing distance, orientation, and predicted travel efficiency could significantly improve overall navigation performance and reduce unnecessary movements.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Directorate General of Vocational Education, Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia, for funding this research through the Master's Thesis Research Grant, as stated in Decree Number 67/SPK/D.D4/PPK.01.APTV/III/2024. The authors also extend their appreciation to the Center of Research and Community Service (P3M) Politeknik Elektronika Negeri Surabaya (PENS) for facilitating the research through Contract Number 516/PL14/PT.01.05/III/2024.

Special thanks are also extended to all members of the Social Robotics and Smart System Applications Laboratory (SRSSA Lab) for their valuable support and collaboration during the study. Finally, the authors gratefully acknowledge the constructive feedback and insightful suggestions provided by the anonymous reviewers, which significantly improved the quality of this manuscript.

REFERENCES

[1] H.I. Krebs and N. Hogan, "Therapeutic Robotics: A Technology Push: Stroke rehabilitation is being aided by robots that guide movement of shoulders and elbows, wrists, hands, arms and ankles to significantly improve recovery of patients," *Proc IEEE Inst Electr Electron Eng.*, vol. 94, no. 9, pp. 1727 – 1738, 2006.

[2] Passler J. Tilley, "Automation, robotics, and the factory of the future," The Great Re-Make: Manufacturing for Modern Times.

[3] D.F. Yépez-Ponce, J.V. Salcedo, P.D. Rosero-Montalvo, and J. Sanchis, "Mobile robotics in smart farming: current trends and applications," Front Artif Intell., vol. 6, pp. 1213330, 2023.

[4] I. Kubasáková, J. Kubánová, D. Benco, and N. Fábryová, ``Application of Autonomous Mobile Robot as a Substitute for Human Factor in Order to Increase Efficiency and Safety in a Company," Applied Sciences, vol. 14, no. 13, pp. 5859, 2024.

[5] A. Tuomi, I.P. Tussyadiah, and J. Stienmetz, "Applications and Implications of Service Robots in Hospitality," Cornell Hospitality Quarterly, vol. 62, no. 2, pp. 232 – 247, 2021.

[6] L. Wijayathunga, A. Rassau, and D. Chai, "Challenges and Solutions for Autonomous Ground Robot Scene Understanding and Navigation in Unstructured Outdoor Environments: A

Review," Applied Sciences, vol. 13, no. 17, pp. 9877, 2023.The Contiki operating system. Available: http://www.contiki-os.org/index.html.

[7] H.M. Mu'allimi, B.S.B. Dewantara, D. Pramadihanto, and B.S. Marta, "Human Partner and Robot Guide Coordination System Under Social Force Model Framework Using Kinect Sensor," The 22nd International Electronics Symposium (IES), 2020.

[8] A.T. Rifqi, B.S.B. Dewantara, D. Pramadihanto, and B.S. Marta, "Fuzzy Social Force Model for Healthcare Robot Navigation and Obstacle Avoidance," 2021 International Electronics Symposium (IES), 2021.

[9] B.S.B. Dewantara and J. Miura, "Generation of a Socially Aware Behavior of a Guide Robot Using Reinforcement Learning," IEEE International Electronics Symposium (IES), 2016.

[10] B.S.B. Dewantara and B.N.D. Ariyadi, "Adaptive Behavior Control for Robot Soccer Navigation Using Fuzzy-based Social Force Model," Smart Science, vol. 9, no. 1, pp. 14 – 29, 2021.

[11] A.C.C. Meng, M. Wand, and V.S.S. Hwang, "A Methodology of Map-Guided Autonomous Navigation with Range Sensor in Dynamic Environment," Proc. SPIE 1007, Mobile Robots III, 1989.

[12] C. Chen and Y. Cheng, "Research on Map Building by Mobile Robots," 2008 Second International Symposium on Intelligent Information Technology Application, Shanghai, China, 2008, pp. 673 – 677, 2008.

[13] J. Achterhold, S. Guttikonda, J.U. Kreber, H. Li, and J. Stueckler, "Learning a Terrain- and Robot-Aware Dynamics Model for Autonomous Mobile Robot Navigation," Cornell University, arXiv:2409.11452, 2024.

[14] B. Yamauchi, "A frontier-based approach for autonomous exploration," Proceedings 1997 IEEE International Symposium on Computational Intelligence in Robotics and Automation CIRA'97, pp. 146 – 151, 1997.

[15] M. Sasaki, Y. Tsuda, and K. Matsushita, "Development of Autonomous Mobile Robot with 3DLidar Self Localization Function using Layout Map," Electronics, vol. 13, pp. 1082, 2024.

[16] A. Ramadhan, B. Dewantara, and S. Setiawardhana, "Optimization of Fuzzy Social Force Model Adaptive Parameter using Genetic Algorithm for Mobile Robot Navigation Control," Jurnal Rekayasa Elektrika, vol. 19, no. 1, 2023, doi: 10.17529/jre.v19i1.28330.

[17] D. Helbing and P. Molnar, "Social force model for pedestrian dynamics," Phys Rev., vol. E51, pp. 4282 – 4286, 1995.

[18] H. Taheri and Z.C. Xia, "SLAM; definition and evolution," Engineering Applications of Artificial Intelligence, vol. 97, 2021.

[19] A.A. Housein and G. Xingyu, "Simultaneous Localization and Mapping using differential drive mobile robot under ROS," Journal of Physics: Conference Series, vol. 1820, 2021 International Conference on Mechanical Engineering, Intelligent Manufacturing and Automation Technology (MEMAT), 2021.

[20] S. Macenski and I. Jambrecic, "SLAM Toolbox: SLAM for the dynamic world," JOSS, vol. 6, no. 61, p. 2783, May 2021.

[21] G. Ferrer et al., "Robot companion: a social-force based approach with human awareness-navigation in crowded environments," IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1688 – 1694, 2013.

[22] F. Zanlungo et al., "Social force model with explicit collision prediction," EPL (Europhysics Letters), vol. 93, pp. 68005-p1 – 68005-p6, 2011.

[23] M. Luber et al., "People tracking with human motion predictions from social forces," IEEE International Conference on Robotics and Automation, pp. 464 – 469, 2010.

[24] N. Chindakham et al., "Simultaneous Calibration of Odometry and Head-Eye Parameters for Mobile Robots with a Pan-Tilt Camera," Sensors, vol. 19, no. 16, p. 3623, Aug. 2019.

[25] M. Keidar and G. A. Kaminka, "Efficient frontier detection for robot exploration," The International Journal of Robotics Research, vol. 33, no. 2, pp. 215 – 236, Feb. 2014.

[26] P. Quin, A. Alempijevic et al., "Expanding wavefront frontier detection: An approach for efficiently detecting frontier cells," Australasian Conference on Robotics and Automation (ACRA), 2014.

[27] Y. Sun and C. Zhang, "Efficient and Safe Robotic Autonomous Environment Exploration Using Integrated Frontier Detection and Multiple Path Evaluation," Remote Sensing, vol. 13, no. 23, p. 4881, Dec. 2021.

[28] ROBOTIS-GIT, "turtlebot3_simulations," GitHub repository}, available at: https://github.com/ROBOTIS-GIT/turtlebot3_simulations.git, accessed on: November 2024.