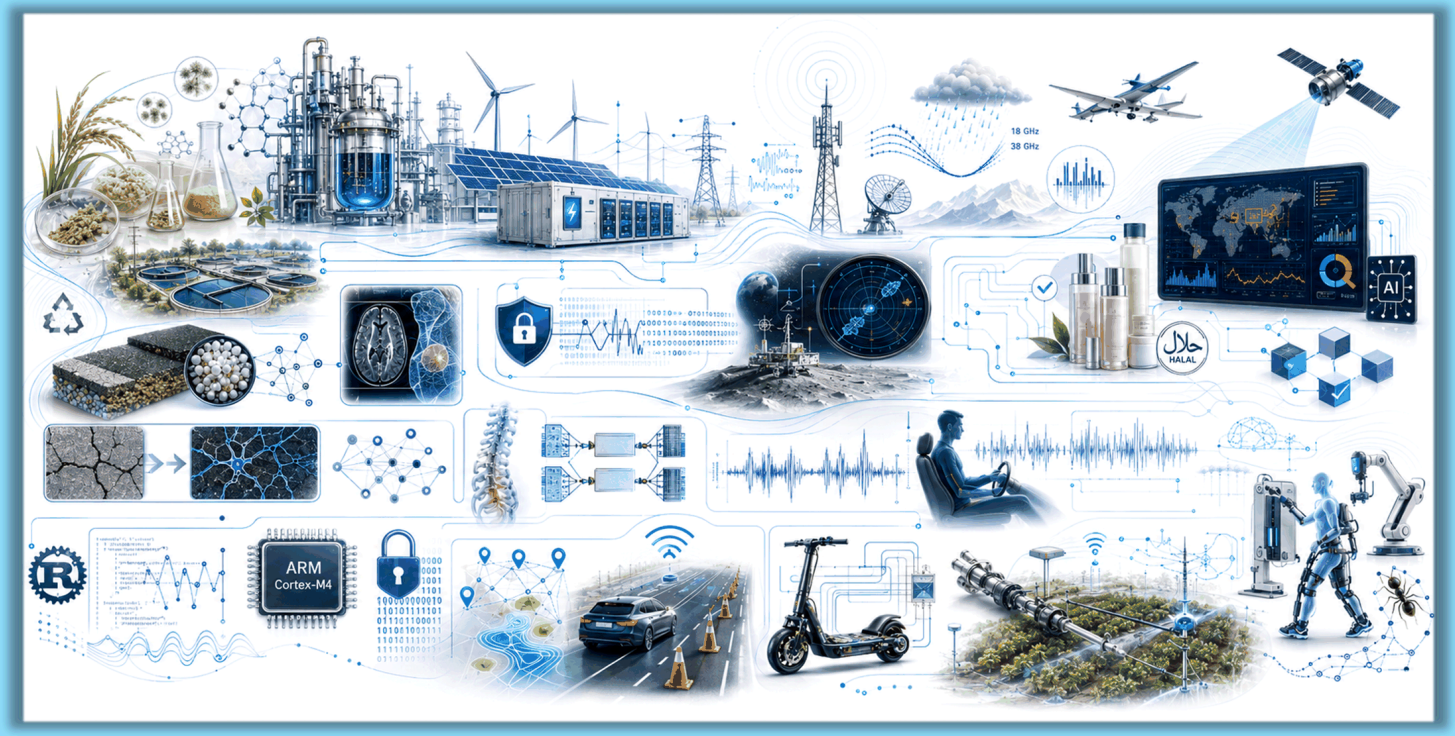
IIUM ENGINEERING JOURNAL

Volume 27 Number 2

May 2026



IIUM Press

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

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

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

IIUM ENGINEERING JOURNAL

Volume 27, Issue 2, May 2026

https:/doi.org/10.31436/iiumej.v27i2

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) hold 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
Abdumalik Rakhimov, IIUM, Malaysia
Ani Liza Asnawi, IIUM, Malaysia
Konstantin Khanin, University of Toronto, Canada
Md Rafiqul Islam, IIUM, Malaysia
Mohamed B. Trabia, United States
Mohammad S. Alam, Texas A&M University-Kingsville, United States
Mohammed Saedi Jami, IIUM, Malaysia
Mohd Hanafi Ani, IIUM, Malaysia
Mohd Sultan Ibrahim Shaik Dawood, IIUM Malaysia
Mustafizur Rahman, National University Singapore, Singapore
Nor Fadhillah Mohamed Azmin, IIUM, Malaysia
Norun Farihah Abdul Malik, IIUM, Malaysia
Ossama Abdulkhalik, Michigan Technological University, United States
Sheroz Khan, Saudi Arabia
Tanveer Saleh, 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.v27i2.4413

Editorial Message Vol. 27 No. 2 2026

Engineering for Resilient, Intelligent, and Sustainable Futures

It is with great pleasure that we present the May 2026 edition of the IIUM Engineering Journal, an issue that reflects the expanding role of engineering in shaping a more sustainable, intelligent, safe, and resilient future. The contributions in this volume demonstrate that engineering research today is no longer confined to solving isolated technical problems. It is increasingly directed toward systems, communities, environments, and technologies that must perform reliably in complex real-world conditions.

A strong sustainability thread runs throughout this issue. Several contributions explore how waste materials, biological resources, renewable energy systems, and intelligent control can be transformed into practical engineering solutions. Studies on microbial bio-coagulant production, sludge remediation, eco-friendly asphalt mixtures, precision irrigation, and energy storage control for renewable microgrids show that sustainable engineering requires more than good intentions. It requires process optimization, material innovation, system-level thinking, careful validation, and the courage to rethink conventional practices.

This issue also highlights the growing maturity of intelligent and data-driven engineering. Artificial intelligence, machine learning, deep learning, foundation models, blockchain, and optimization techniques are applied across diverse problems, including medical image classification, spinal disorder diagnosis, halal verification, satellite anomaly identification, crack segmentation, intrusion detection, facility location, and autonomous navigation. These works show a clear movement beyond algorithmic novelty toward explainability, robustness, efficiency, and practical deployment. In this sense, intelligence is not treated merely as computational performance, but as a means to support better decisions, safer systems, and more trustworthy engineering outcomes.

Equally important is the emphasis on safety, resilience, and trust. The papers on communication reliability, software-defined radar, tropical rain attenuation, embedded cryptography, vehicle safety alerts, highway cone deployment, UAV dynamics, electric vehicle control, and lunar navigation demonstrate the importance of engineering systems that can withstand uncertainty, disturbance, failure, and demanding operating environments. These studies remind us that high-performance engineering is incomplete without reliability, accountability, and resilience.

The human dimension of engineering is also clearly visible in this volume. Contributions on rehabilitation robotics, EMG-based assessment during prolonged driving, steering control, mobility systems, biomedical classification, and Rust-based visualization for teaching concurrency demonstrate that engineering research ultimately serves people. It supports health, safety, learning, mobility, productivity, and quality of life. This human-centered direction is especially important as engineered systems become more autonomous, more digital, and more deeply embedded in daily life.

Taken together, this May 2026 edition presents a compelling picture of engineering as purposeful innovation. The articles are diverse in method, domain, and application, yet they share a common commitment to transforming knowledge into responsible solutions. They also reflect the kind of scholarship that the IIUM Engineering Journal seeks to promote: rigorous in method, relevant in application, interdisciplinary in spirit, and meaningful in contribution.

On behalf of the editorial team, we sincerely thank the authors for their valuable contributions, the reviewers for their careful and constructive evaluations, and the editorial board members for their dedication to maintaining the journal's quality and integrity. We hope this issue will inspire further research, strengthen collaboration across disciplines, and encourage engineering innovations that contribute to society and the world in a wise and responsible way.

Prof. Ir. Ts. Dr. Teddy Surya Gunawan

Executive Editor

IIUM Engineering Journal

IIUM ENGINEERING JOURNAL

Volume 27, Issue 2, May 2026
https:/doi.org/10.31436/iiumej.v27i2

Table of Contents

Copyright i
Editorial Board ii
Editorial Message iv

Chemical and Biotechnology Engineering

4154: Solid-state Fermentation for Production of Bio-coagulant on Rice Bran using Candida utilis, Trichoderma sp., Phanerochaete chrysosporium for Turbidity Removal 1
Radhia Nedjai, Abdullah Al Mamun, Md Zahangir Alam

4237: Optimization and Scale-Up of Solid-State Bioconversion for Microbial Coagulant Production toward Sustainable Water Treatment 15
Maroua Fellah, Md Zahangir Alam, Abdullah Al Mamun, Abdelmohsen Benoudjit, Nassereldeen Ahmed Kabbashi, Nurul Sakinah Engliman

Civil and Environmental Engineering

3843: Eco-Friendly Asphalt Mixture: Utilization of Waste Polypropylene as a Partial Replacement of Aggregate 31
Israa Daryul, Roaa Latief

3858: Enhancement of Stabilization/Solidification Method with Bioaugmentation for Sludge Remediation 51
Muhammad Hasanuddin Mohd Kamal, Nasuha Mohd Rais, Nor Amani Filzah Mohd Kamil, Nur Fateha Md Nor

Electrical, Computer and Communications Engineering

3387: Magnetic Resonance Imaging and Ultrasound-based Breast Tumor Classification using Soft Voting Ensemble Deep Learning Method 74
Suresh Ramayanam, Vijaya Bathini, Sankar Babu Jangili, Santosh Kumar Reddy Thikkam

3683: New Control Strategy of Battery Energy Storage System in Microgrids Integrated with Wind and Solar Power Plants 87
Duy Quang Nguyen, Minh Khoa Ngo

3951: Comparative Performance Analysis of Forward Error Correction Schemes in AX.25-Based APRS Communication Systems with AFSK Modulation using GNU Radio Simulation 117
Rifki Ardinal, Agfianto Eko Putra, Wahyudi Hasbi, Eriko Nasemudin Nasser, Widya Roza

3969: Quantitative Assessment of Coffee Quality Using Microwave Non-Destructive Testing
Mohd Khairil Adzhar Mahmood, Zuhani Ismail Khan, Megat Syahirul Amin Megat Ali, Ahmad Ihsan Mohd Yassin, Norziaton Ismail Khan, Michael Kwan Yeow Tan 129

3988: Experimental Validation of a Labview-Based SFCW Software-Defined Radar System .
Nor Hazima Ardzemi, Farah Nadia Mohd Isa, Norun Farihah Abdul Malek, Huda Adibah Mohd Ramli, Ahmad Fadzil Ismail 146

4008: Raindrop Size Distribution and Wind-Driven Rain Specific Attenuation: CCDF-Based Analysis at 18 and 38 GHz in Tropical Country 162
Nurul Najwa Md Yusof, Omar Abdul Aziz, Lam Hong Yin, Jafri Din

4059: MultiModal Explainable AI and Blockchain Integration for Automated Halal Verification of Cosmetic Products 175
Endang Supriyati, Mohammad Iqbal, Tri Listyorini

4061: Generative AI Models: A Comparison of Application Analysis on Web AI-Based Decision Support Systems for Satellite Anomaly Identification..... 189
Abdul Mutholib, Nadirah Abdul Rahim, Teddy Surya Gunawan, Ahmad Shah Hizam Md Yasir

4124: PromptLessSAM: From Foundational Model to Domain Expert via Lightweight Decoder Adaptation for Crack Segmentation 210
Mohammed Al-Mustafa Hendo, Israa H. Ali

4240: Stability-Aware Evaluation of a CNN–LSTM–DQN Intrusion Detection System for Zero-Day and Drifted Network Traffic..... 227
Rushendra, Kalamullah Ramli, Prima Dewi Purnamasari

4246: A Robust Low-Complexity Star Centroiding Algorithm for Autonomous Navigation under Lunar Noise Conditions 257
Anisa Hannani Razaman, Yasser Asrul Ahmad, Teddy Surya Gunawan, Othman Omran Khalifa, Mila Popović

4341: Efficiency-Aware Multi-Class Spinal Disorder Classification Using CBAM-Enhanced Lightweight CNNs with Dual-Branch Fusion..... 278
Teddy Surya Gunawan, Nurul Jannah, Mira Kartiwi, Noreha Abdul Malik

4361: TSGL-Rust: Teaching Concurrency Through Compile-Time Enforcement via Rust-Based Visualization of Classical Synchronization Problems..... 320
Teddy Surya Gunawan

4405: Constant-Time Bitsliced Rijndael-256 on ARM Cortex-M4: On the Limitations of Fixslicing Beyond AES-128 340
Andriani Adi Lestari, Suryadi MT, Kalamullah Ramli, Teddy Surya Gunawan, Esti Rahmawati Agustina, Susila Windarta

Engineering Mathematics and Applied Science

4072: Scalable Multi-Objective Optimization for Facility Location Using A Metaheuristic Technique..... 363
Thanathorn Karot, Choosak Pornsing

Mechanical and Aerospace Engineering

3968: Optimum Alert For In-Vehicle Safety Technologies..... 377
Noor Irza Mohd Zaki, Mohd Khairi Abu Husain, Nurulakmar Abu Husain, Siti Mahirah Che Husin, Yahaya Ahmad

3971: Field and Laboratory Evaluation of Reliability in a Novel Cone Deployment System for Safer Highway Operations 397
Azilah Ismail, Mohd Azli Salim, Nor Azmmi Masripan, Mohd Zaid Akop, Siti Wardatulaina Mohd Yusof, Chonlatee Photong, Alan Watson, Mohammed Hussin A. Al-Mola

4013: Analytical and Flight Test-Based Dynamic Characteristic Identification of a Small Fixed-Wing UAV..... 415
Eries Bagita Jayanti, Yazdi Ibrahim Jenie, Ardian Rizaldi, Novita Atmasari, Fuad Surastyo Pranoto

4062: Impact of Aggressive and Moderate Driving Intensity on Vehicle Air–Fuel Ratio Stability 434
Mohamad Norsyafiq Ab Hamid, Muhammad Yusri Bin Ismail, Najmi Haziq Badrulhisam

4068: Design Optimization and Structural Evaluation of a Hybrid Frame for Next-Generation Electric Scooters 450
Sridhar R, Mahantesh M Math, Sujan Chakraborty, Sourabha H, Vivekanand S Gog, Jeevathith R

Mechatronics and Automation Engineering

3647: A Robust Sensor-Failure-Tolerant Fuzzy Control Framework with Predictive Data Imputation for Sustainable Precision Irrigation 467
Indra Laksmana, Swivano Agmal, Rosda Syelly, Melda Amalia, Perdana Putera

3982: Steering Control Precision in Rack Steering Vehicles Using Finite-Time Prescribed Performance Anti-Windup PI Controller with Nonlinear Error Transformation 479
Addie Irawan, Norsharimie Mat Adam, R.M.T Raja Ismail, Aliza Che Amran, Mohd Iskandar Putra Azahar

3991: Quantifying Activity of the Erector Spinae Longissimus Subgroup Using EMG During Prolonged Driving 496
Muhammad Alfatih Eltayeb, Nur Liyana Azmi, Ahmad Jazlan

4021: Integration of Neutrosophic Methods into Adaptive Control of Nonlinear Systems Using Neuro-Fuzzy Networks With B-Spline Functions 506
Oksana Porubay, Dilnoza Umurzakova

4134: A Comprehensive Framework of Ant Colony Optimisation for Optimal Control of Upper Extremity Rehabilitation Robot 529
Muhammad Kamran Joyo, Izanoordina Ahmad, Kushsairy Abdul Kadir, Talha Ahmed Khan, Noaman M Noaman

4136: Optimal and Robust Speed Control of Electric Vehicles using a Hybrid Ant Colony Optimization Tuned Sliding Mode Control 545
Parthiban S, Krishna S

Blank page.

https://doi.org/10.31436/iiumej.v27i2.4154

Solid-state Fermentation for Production of Bio-coagulant on Rice Bran using *Candida utilis, Trichoderma* sp., *Phanerochaete chrysosporium* for Turbidity Removal

RADHIA NEDJAI[1,2*], MD ZAHANGIR ALAM[1], ABDULLAH AL MAMUN[3*]

[1]*Department of Chemical Engineering & Sustainability, Faculty of Engineering, International Islamic University Malaysia (IIUM), Kuala Lumpur, Malaysia.*
[2]*Laboratory of Environmental Engineering, Department of Process Engineering, Badji Mokhtar University, Annaba, 23000, Algeria.*
[3]*Department of Civil Engineering, Kulliyyah of Engineering, International Islamic University Malaysia (IIUM), Kuala Lumpur, Malaysia.*

**Corresponding authors: mamun@iium.edu.my, radhia.nedjai@univ-annaba.dz*

(Received: 24 November 2025; Accepted: 9 January 2026; Published online: 10 May 2026)

ABSTRACT: Solid-state fermentation is a microbial culture technique used to produce compounds and products of industrial interest. In this study, rice bran, a by-product of the rice processing industry, was used as a coagulant substrate for *Candida utilis*, *Trichoderma* sp., and *Phanerochaete chrysosporium* in solid-state fermentation (SSF). To the best of our knowledge, no study has reported the production of coagulants by *Trichoderma* sp. *and Candida utilis*. Moreover, no attempts have been made to utilize rice bran as a substrate for the production of coagulants in solid-state by these three microorganisms. Two different substrate proportions were used to produce natural coagulants. Bio-coagulants produced in this research were used to remove turbidity from the kaolin solution. The experiments were conducted with artificially turbid water and a conventional jar test apparatus. The findings demonstrate that Substrate 2, distinguished by its lower water content, exhibited the highest turbidity removal efficiency compared to Substrate 1, which possessed higher water content. Moreover, *Trichoderma* sp. showed a maximum turbidity removal of 71%. While the maximum turbidity removal by the coagulants produced from *C. utilis* and *P. chrysosporium* was 59 %. The microscopic observation of the microorganism's growth on rice bran substrate has also been performed. It was revealed that the three microorganisms fully colonized the substrate surface. Further optimization of the SSF process parameters is expected to achieve efficient turbidity reduction in wastewater.

ABSTRAK: Penapaian keadaan pepejal ialah teknik kultur mikrob yang digunakan untuk menghasilkan sebatian dan produk yang mempunyai kepentingan industri. Dalam kajian ini, dedak padi, iaitu hasil sampingan daripada industri pemprosesan beras, digunakan sebagai substrat koagulan bagi *Candida utilis*, *Trichoderma* sp., dan *Phanerochaete chrysosporium* dalam penapaian keadaan pepejal (*solid-state fermentation*, SSF). Sepanjang pengetahuan kami, belum terdapat kajian yang melaporkan penghasilan koagulan oleh *Trichoderma* sp. dan *Candida utilis*. Selain itu, belum ada usaha yang dilaporkan untuk memanfaatkan dedak padi sebagai substrat bagi penghasilan koagulan secara keadaan pepejal oleh ketiga-tiga mikroorganisma ini. Dua nisbah substrat yang berbeza telah digunakan untuk menghasilkan koagulan semula jadi. Bio-koagulan yang dihasilkan dalam kajian ini digunakan untuk menyingkirkan kekeruhan daripada larutan kaolin. Eksperimen dijalankan menggunakan air keruh buatan dan radas ujian balang konvensional. Dapatan kajian menunjukkan bahawa Substrat 2, yang dicirikan oleh kandungan air yang lebih rendah, mempamerkan kecekapan penyingkiran kekeruhan yang paling tinggi berbanding Substrat 1 yang mempunyai kandungan air lebih tinggi. Selain itu, *Trichoderma* sp. menunjukkan penyingkiran kekeruhan

maksimum sebanyak 71%, manakala penyingkiran kekeruhan maksimum oleh koagulan yang dihasilkan daripada *C. utilis* dan *P. chrysosporium* ialah 59%. Pemerhatian mikroskopik terhadap pertumbuhan mikroorganisma pada substrat dedak padi juga telah dijalankan, dan menunjukkan bahawa ketiga-tiga mikroorganisma tersebut telah mengkolonikan permukaan substrat sepenuhnya. Pengoptimuman lanjut terhadap parameter proses SSF dijangka dapat mencapai pengurangan kekeruhan yang lebih berkesan dalam air sisa.

KEYWORDS: *Bio-Coagulant, Rice bran substrate, Large-scale, Solid-state fermentation, Turbidity removal.*

1. INTRODUCTION

Chemical flocculants are frequently used in downstream processing, the food, fermentation, wastewater, and drinking treatment industries, owing to their great flocculating efficiency and cost-effectiveness. However, because of their widespread use, there are significant environmental and health risks [1]. Therefore, biotechnology research has focused on developing more environmentally friendly and safe coagulants for pollutant removal in water treatment to eliminate the hazards associated with chemical coagulants. Thus, bio-coagulants have recently attracted global attention due to their enormous potential to replace chemical flocculants [2].

Bio-coagulants are mainly polysaccharides and proteins [3] that can be isolated from plant, animal, or microbial tissues [2]. Because of their biodegradability, they are more environmentally benign than both organic and inorganic coagulants [3]. Bio-coagulants provide several advantages over chemical coagulants, including minimal toxicity, low residual sludge generation, and low cost [4]. These qualities make bio-coagulants useful for treating water and as potential substitutes for chemical coagulants. However, compared to synthetic organic flocculants, bio-coagulants have limited applicability in water treatment because they require sophisticated extraction techniques and high operational costs. Accordingly, it is necessary to have a coagulant/flocculant with low operating expenses and high flocculation efficiency [5].

Solid-state fermentation is a fungal or bacterial culture technique used to produce compounds and products of industrial interest [6], including pre-treatment of agricultural wastes for producing simple sugars (e.g., glucose) [7], antibiotics (e.g., griseofulvin) [8], pigments [9], enzymes (e.g., cellulases) [10], flavor and fragrance compounds [11], organic acids [12] and organic compounds (e.g., surfactants) [13]. In SSF, the required microorganisms are cultivated on an inert or natural substrate as a solid support, with minimal or no free water [14]. Many researchers have found that using SSF has various biotechnological benefits, including higher fermentation productivity, higher product end concentrations, higher product stability, lower catabolic repression, cultivation of microorganisms specialized for water-insoluble substrates or mixed cultivation of diverse fungi, and finally, fewer sterility demands owing to the low water activity used in SSF [9]. Furthermore, SSF provides many benefits over conventional submerged fermentation (SF), including a simpler process, lower energy use, less pollution, greater product recovery, and a more closely simulated native microorganism environment [14].

Typically, homogeneous agricultural products or lignocellulosic wastes are excellent substrates for SSF because they contain the carbon, nitrogen, and various mineral elements (K, Mg) necessary for microbial growth and metabolite synthesis [14]. However, only a few studies have used solid-state fermentation (SSF) to produce bio-coagulants. Luthfi et al. (2020) used SSF to produce a bioflocculant from Aspergillus niger DWB, employing oil palm empty fruit

bunch fiber as the substrate [15]. Mamun et al. (2023) and Nedjai et al. (2024) employed SSF with coco peat as the substrate to produce a myco coagulant [16,17]. Nedjai et al. (2025) also used coco peat as the substrate to produce a powdered myco coagulant from *Phanerochaete concrescens* [18]. A variety of substrates have been used for the growth of fungi, including seeds, roots, rice, wheat bran, and wood pieces. Several factors affect substrate selection, but the two most crucial ones are availability and cost [19].

Rice bran is produced as a byproduct of milling during the processing of paddy. Rice is the second-most widely grown crop worldwide, particularly in developing countries. Milling of paddy yields 8% rice bran [20], with one tonne of milled rice yielding 60-80 kg of rice bran [21]. Despite the enormous volume of rice bran produced, it is still not used enough for human consumption [20]. Thus, the use of rice bran as a substrate is an efficient solution to the disposal of large amounts of waste and serves as a waste-to-wealth resource, as leftover rice bran is used to produce a beneficial coagulant that could help many countries address their wastewater treatment crises.

Consequently, the main objective of this research is to use rice bran to produce coagulants via SSF with Candida utilis, Trichoderma sp., and Phanerochaete chrysosporium, using the tray method for semi-industrial production. To the best of our knowledge, no research has reported on the production of coagulants by *Trichoderma sp.* and *Candida utilis*. Furthermore, no attempts have been made to use rice bran as a substrate for producing coagulants via solid-state fermentation by these three microorganisms. Natural coagulants were produced using two distinct substrate proportions. Natural coagulants were evaluated for turbidity removal from a kaolin solution. So, the experiments were carried out using artificially turbid water and a standard jar test apparatus. Scanning electron microscopy (SEM) was also used to observe the growth of the three microorganisms on the rice bran substrate.

2. MATERIALS AND METHODS

2.1. Materials

The fungal strains (*Trichoderma* sp. and *Phanerochaete chrysosporium*) and the bacterial strain (*Candida utilis*) were obtained from the Department of Chemical Engineering and Sustainability, Kulliyyah of Engineering, International Islamic University Malaysia (IIUM). The microbial strains were transferred into a refrigerator and preserved on Potato Dextrose Agar (PDA) at 4°C for future use. Oxoid Ltd. provided the Malt Extract Broth (MEB). We bought potato dextrose agar (PDA) from Difco™. R&M Chemicals provided NaOH, kaolin, D(+)-glucose, and a buffer solution at pH 7. The rice bran was crushed and sieved to obtain fine powders that passed through a 50 mm mesh.

2.2. Methods

2.2.1. Preparation of Media

In this investigation, MEB and glucose were employed. 3% MEB and 2% D (+)-glucose were combined with distilled water to create 100 mL of medium. After adjusting the mixture's pH to 6 with NaOH using a pH meter (METTLER TOLEDO, PB-10 Sartorius, Germany), it was combined with rice bran and autoclaved for 15 minutes at 121°C in an autoclave (Hirayama, HV-110, HVE-50).

2.2.2. Microbes Isolation and Cultures

20 mL of PDA medium, prepared according to the manufacturer's instructions, was used to culture microbes in Petri dishes. To ensure that everything had completely dissolved, the

PDA was suspended in 1 L of distilled water, heated for about 60 s, and then vigorously stirred using a hot plate (Jenway Ltd., UK, Model 1000). Before being placed in sterilized Petri dishes, the mixed components were autoclaved at 121°C for 15 minutes and allowed to polymerize for at least 30 minutes. Microbe strains were inoculated. The yeast *Candida utilis* and *the basidiomycete Phanerochaete chrysosporium plates were incubated at 30 °C for 2-3 days [7] to achieve* sufficient growth. The Trichoderma sp. plates were incubated at 30 °C for 5 days [21] to complete mycelial growth. The isolated microbes were purified and further inoculated onto PDA plates in a biological safety cabinet (ESCO Technologies Inc., USA).

2.2.3. Preparation of Kaolin Solution

In this experiment, kaolin was used to replace water. A specific quantity of kaolin powder was dissolved in 10 liters of distilled water to prepare the stock kaolin solution with an initial turbidity of 750 ± 10 NTU. A Jar test device (Model SW6, Bibby Stuart, UK) was used to gently mix the liquid at 200 rpm for 1 h to achieve a uniform dispersion of kaolin particles. Afterward, the suspension was allowed to stand for a day to ensure that the kaolin was completely hydrated [16].

2.2.4. Cell and Spore Suspensions Preparation

The suspensions of *Candida utilis* cells, *Trichoderma* sp., and *P. chrysosporium* spores were prepared from their cultures. The yeast cells of *Candida utilis* were washed by suspending them in sterile distilled water and then collected using a micropipette. In contrast, the fungi were collected by scraping the media on the plates with a sterile loop and then transferring them to a beaker filled with sterile distilled water. The breakers were shaken using shakers (SK71 model, Jeio Tech Co., Ltd., Korea) to separate spore aggregates. The volume of distilled water was subtracted from the 18% inoculum volume used [18].

2.2.5. Production of Coagulants

To produce the coagulant, rice bran was mixed with an additional amount of media. There were two different proportions used. In Substrate 1, 41.25 g of rice bran and 124.87 mL of media were combined, while in Substrate 2, 83.25 g of rice bran and 83.25 mL of media were combined. The mixture was poured into aluminum trays and covered with aluminum foil. The mixes were then autoclaved for 15 minutes at 121ºC. Afterward, the mixtures were poured into the sterilized trays. The trays were inoculated with suspensions of *Candida utilis* cells, *Trichoderma* sp., and *P. chrysosporium* spores to promote additional inoculum expansion. After covering the trays with aluminum foil, they were incubated at 30 °C for 10 days using a laboratory incubator (Memmert Model).

2.2.6. Coagulants Extraction

4 g of each culture was placed into conical flasks and combined with 20 mL of buffer solution after various days of incubation (1-10 days) (pH 7). At room temperature and 250 rpm, the flasks were agitated for 1 hour. The coagulant was separated from the biomass using a centrifuge (Hettich® Universal 320R, Germany) at 9000 rpm for 5 minutes at 25 ºC. The supernatants were collected, and their flocculating activity was determined by the Jar apparatus method [17].

2.2.7. Evaluation of Produced Coagulants

The kaolin clay suspension method was used to measure the flocculation activity. The initial and final turbidity were measured using a turbidity meter (EXTECH Instruments, TB400, China). In separate beakers containing 300 mL of kaolin solution (750 ± 10 NTU), 10 mL of coagulant was added. After that, the beakers were then shaken for several durations at

different rates, including 7 minutes of rapid mixing at 250 rpm and 22 minutes of slow mixing at 90 rpm. The suspensions were allowed to settle for 60 minutes after agitation was stopped. Using a pipette, samples were taken from the center of the supernatant to determine the final concentration.

The flocculation activity was determined using the following equation:

$$Turbidity\ efficiency\ (\%) = \frac{(initial\ turbidity - final\ turbidity)}{final\ turbidity} \times 100 \qquad (1)$$

2.2.8. Statistical Analysis

All analyses were performed in triplicate. Graphs and statistical analyses were carried out using Microsoft Excel (version 16.69.1). One-way ANOVA was used to analyze the turbidity removal results. Statistically significant differences among microbes were determined by Tukey's test.

2.2.9. Scanning Electron Microscopy (SEM)

A JEOL-IT 100 SEM device was used to examine the morphology and microstructure of microorganisms on rice bran. The morphology of the isolates' colonies was examined after 3 days of incubation. All samples were placed on carbon tape, transferred to a metallic stub, and then allowed to dry in air under a hood. Afterward, using a Quorum Q300TD with a sputter current of 20 mA, the samples were sputter-coated with a thin layer of gold (10–20 nm) for 50 seconds. Cell size measurements were performed using the measurement tool on the JEOL IT100 SEM (InTouchScope^{TM}).

3. RESULTS AND DISCUSSION

3.1. Morphological Structure of Microorganisms

After the incubation period, the plates were examined, and microscopic observation was performed (Figure 1).

Figure 1 (a) shows *C. utilis* cultivated on PDA media for 3 days at 30ºC. On PDA, *C. utilis* appears as smooth, white-to-cream colonies. They had an entire margin and a convex elevation. A microscopic structure of *C. utilis is* presented in Figures 1(b) and 1(c). *C. utilis* yeasts were single, unicellular, oval cells. Some budding (not separated) cells were also observed.

Trichoderma sp. was grown on PDA media for 5 days at 30 °C (Figure 1(d). After 24 hours of incubation at 30°C, conidial germination was first observed, and the colony reached a radius of 45 mm after 5 days. Cottony white at first; eventually, the colony changed to light green, then dark green, as sporulation spread over the agar plate during the incubation period. Microscopic observation of *Trichoderma* sp. showed the presence of dark green conidial masses (Figure 1 (e)). The conidia of *Trichoderma* sp. were subglobose to ovoidal (Figure 1 (f)). The chlamydospores were not observed.

Figure 1 (g) shows the growth of *P. chrysosporium* on PDA media for three days at 30ºC when the mycelial mats filled the Petri dish. The *P. chrysosporium* colony structures were thin, white, cottony, and sticky. A microscopic observation of *P. chrysosporium* showed unbranched, smooth-walled hyphae (Figure 1(h), 1(i)). Chlamydospores were also observed (Figure 1(i)).

The rapid growth and fluffy appearance of the microorganisms' colonies on PDA provide clues about potential bio-coagulant production, which is efficient in removing turbidity from water.

https://doi.org/10.31436/iiumej.v27i2.4154

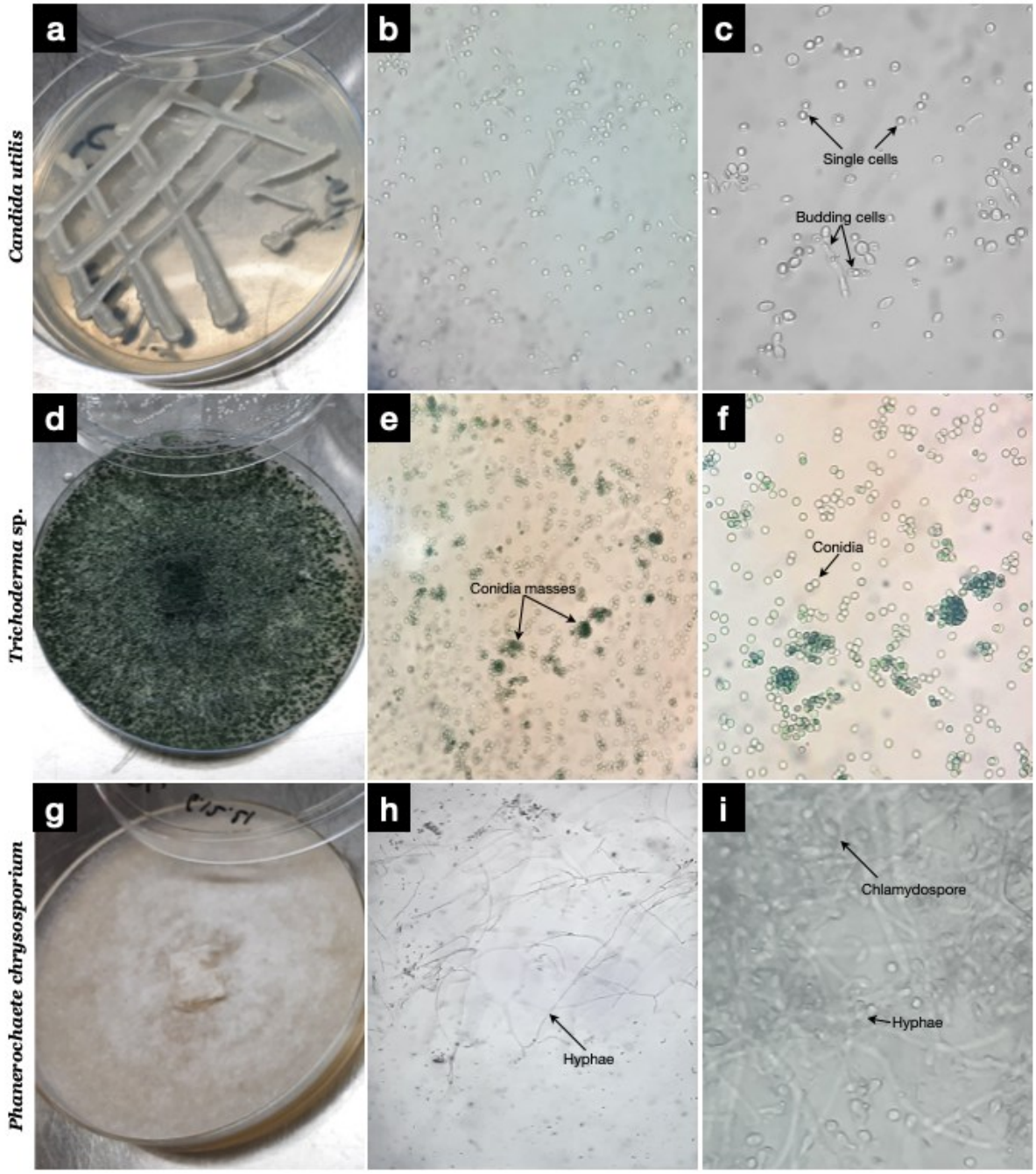


Figure 1. Macroscopic and microscopic features of yeast cells of *Candida utilis*: (a) Colony appearance on PDA plate, (b) magnification ×10, (c) magnification ×40; *Trichoderma* sp.: (d) Colony appearance on PDA plate, (e) magnification ×10, (f) Conidia magnification ×40; and *Phanerochaete chrysosporium*: (g) Colony appearance on PDA plate, (h) magnification ×10, (i) magnification ×40.

3.2. Microorganisms Growth on the Substrates

Three distinct microorganisms and two substrate ratios were used to produce coagulants, with the aim of determining the optimal growth conditions using the tray method. Microbial cultures on the rice bran after different days (3, 4, and 5 days) are shown in Figure 2. Within 5 days, the microorganisms grew and colonized the entire surface of the substrates, as observed. The color of the surface substrates changed over these three days due to an increase in intensity

caused by microorganism growth. Moreover, the microorganism's growth on the substrate prepared by Substrate 2 (Figure 3) was much higher than that on Substrate 1 (Figure 2).

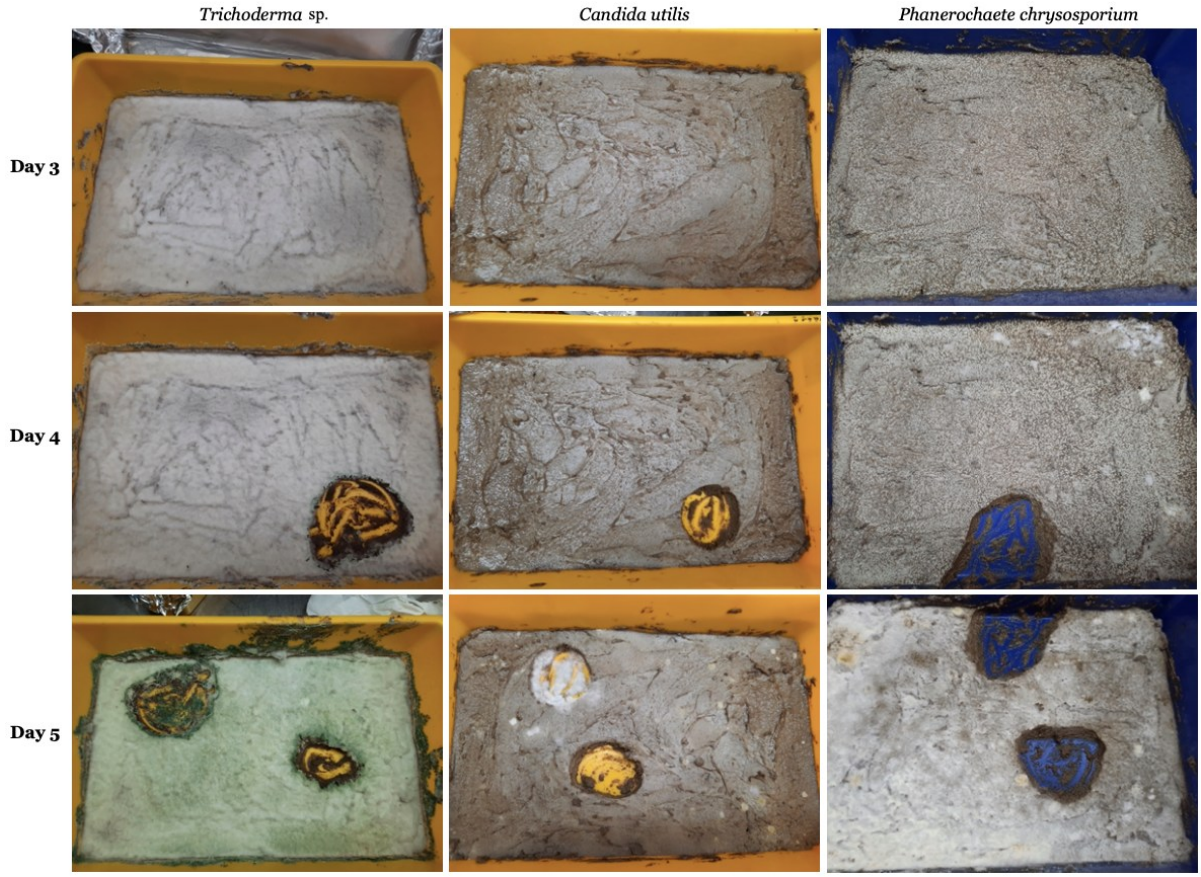


Figure 2. Microbe cultures on substrate 1 after different days (3, 4, and 5 days).

The composition of the substrate influences the concentration of extracted bio-coagulants. Substrates prepared with 83.25 mL of media and 83.25 g of rice bran yielded higher concentrations of bio-coagulants than those prepared with 124.87 mL of media and 41.62 g of rice bran. This difference is attributed to the higher water content present in substrates containing 124.87 mL of media and 41.62 g of rice bran. Excess water content beyond the saturation level that the solid substrate can absorb would reduce interparticle air space and impede oxygen transfer into the SSF [22]. Moisture is generally required for fungi to thrive. However, an oversaturated environment can hinder the life cycle of certain fungi. Rintaro Hattori et al. (2014) reported that excess water inhibits fungal spore germination and hyphal growth [23]. It can be seen that growth on the substrate surface is usually higher due to increased oxygen availability [24].

https://doi.org/10.31436/iiumej.v27i2.4154

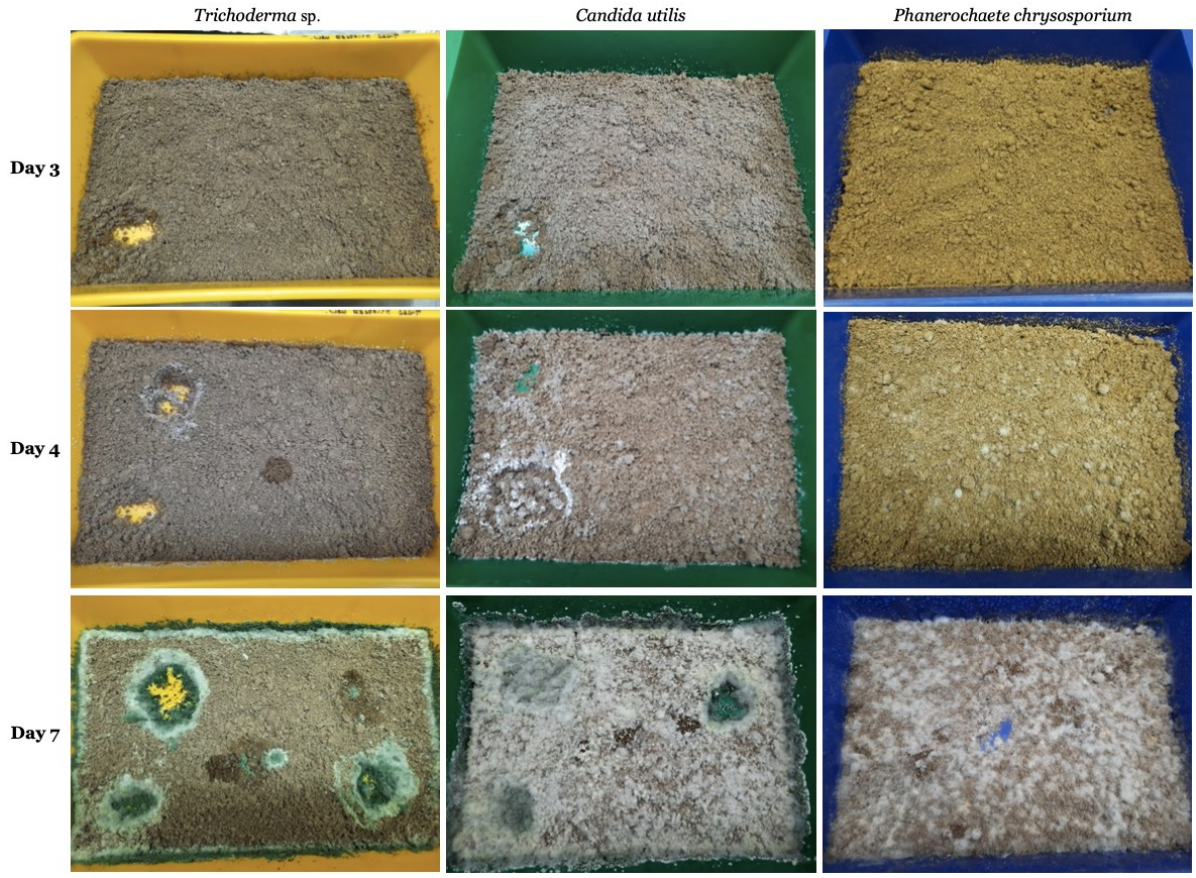


Figure 3. Microbe cultures on Substrate 2 after different days (3, 4, and 5 days).

3.3. Assessment of Turbidity Removal

Jar testing, the primary technique for assessing coagulant efficiency, was employed to evaluate the coagulants produced by three microorganisms in terms of turbidity reduction through flocculation of the kaolin suspension.

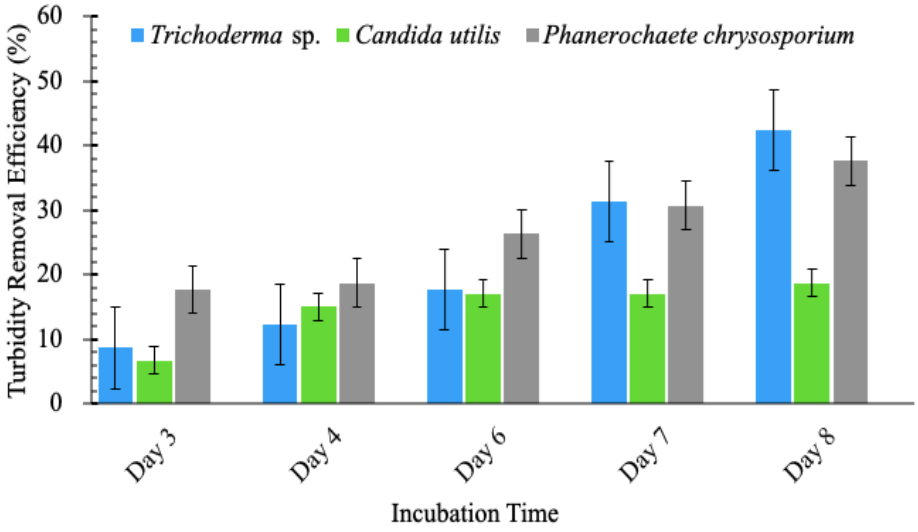


Figure 4. Turbidity Removal (%) from Kaolin Suspension (Substrate 1).

Figures 4 and 5 illustrate the turbidity removal efficiency for different coagulants produced by various microorganisms. Figure 4 presents the turbidity removal of natural coagulants

produced on Substrate 1. Generally, it can be seen that turbidity removal increased progressively each day (from day 3 to day 8), for all produced coagulants. All coagulants achieved their highest turbidity removal on day 8. The turbidity removal by all coagulants produced from *Trichoderma* sp., *C. utilis*, and *P. chrysosporium* achieved their highest turbidity removal of 42%, 19%, and 32%, respectively, on day 8. The most effective coagulant produced in Substrate 1 is produced by *Trichoderma* sp.

From Figure 4, one-way ANOVA gives p =0.283 (p>0.05). The results show no statistically significant differences in turbidity removal efficiency among the three microbes. Tukey HSD shows no pairwise differences: *Trichoderma* sp. and *C. utilis* (p=0.550), *Trichoderma* sp. and *P. chrysosporium* (p=0.832), *C. utilis* and *P. chrysosporium* (p=0.263). All strains perform comparably for bio-coagulant screening.

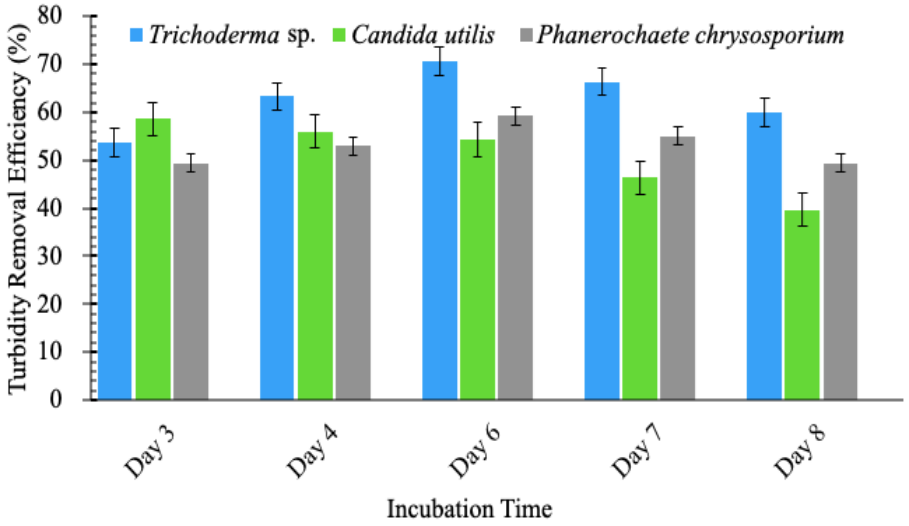


Figure 5. Turbidity Removal (%) from Kaolin Suspension (Substrate 2).

The turbidity removal of natural coagulants produced on Substrate 2 is shown in Figure 5. The findings demonstrated that the turbidity removal of *C. utilis* coagulant increased over the first three days, and this coagulant reached the maximum removal percentage of 59% on day 3. In contrast, turbidity removal by coagulants produced from Trichoderma sp. and *P. chrysosporium* continued to increase until day 6, reaching 71% and 59%, respectively, before declining to 60% and 49% by day 8, respectively. These findings are consistent with Nedjai *et al.* [17], who observed peak myco-coagulant activity on day 6. Dharsana & Prakash (2023) reported that, for a modified banana peel bio-coagulant, approximately 71% turbidity removal was obtained at the highest tested dose (around 10 wt.%) [25]. *C. utilis* and *P. chrysosporium* coagulants achieve moderate turbidity removal (~60%), comparable with several plant-derived and fungal bio-coagulants, such as orange-peel powder (61.63%) [26] and *Aspergillus niger* (60.22%) [27].

Comparing the turbidity removal results of the produced coagulants on two substrates, Substrate 1 showed the lowest turbidity removal for the three microbes. *Trichoderma* sp. performed best on Substrate 2 (71% max), followed by *P. chrysosporium* and *C. utilis* (both 59%). Subsequently, solid-state fermentation with low water content is more advantageous than that with high water content for producing greater numbers of microorganisms, thereby yielding a high concentration of coagulants. This is due to water being a limiting factor for fungal growth in SSF [19,28]. On the other hand, oxygen is usually the major growth-limiting factor in aerobic fermentations of *C. utilis* yeast [29] because oxygen has low solubility in

water [30,31]. Therefore, increasing the water content in the solid substrate reduces yeast growth, thereby producing a low concentration of coagulants.

Coagulation activity declined after several days of microbial growth (day 3 for *C. utilis*; day 6 for fungi; Figure 5), potentially due to dosage plateaus or excess coagulant-restabilizing particles [17]. Alternatively, microbial enzymes may degrade bio-coagulants [32]. Relevant studies indicate that increasing the amount of coagulants/flocculants yields negligible improvements, or even reductions, in turbidity removal efficiency. Excess coagulant fails to interact effectively with suspended particles; instead, it adds additional colloidal solids, intensifying turbidity [17]. Another hypothesis posits that the coagulation activity post-peak might be due to bio-coagulant-degrading enzymes produced by microorganisms [32]. The three microbes exhibited distinct peak turbidity removal times (*C. utilis* on day 3*, Trichoderma* sp*., and P. chrysosporium* on day 6), reflecting species-specific differences in growth and metabolic kinetics and indicating that the production-degradation balance of their bio-coagulants is optimal at different cultivation ages.

All microbes produce extracellular polymeric substances (EPS) that act as bio-coagulants. *C. utilis* mainly produces glucomannan-type EPS [33]. *P. chrysosporium* secretes both soluble EPS (SEPS) and bound EPS (BEPS), with BEPS characterized by a high polysaccharide content [34]. *Trichoderma* sp. similarly synthesizes extracellular polysaccharides that contribute to its biocoagulant activity [35]. The substrate composition, using rice bran supplemented with glucose and malt extract, significantly influences the EPS composition produced by each microbe (monosaccharide profile, protein/polysaccharide ratio, functional groups). The differing EPS compositions of the three microbes help explain the performance differences in turbidity removal observed in Figures 4 and 5.

Trichoderm sp., a versatile lignocellulolytic fungus, likely generated diverse anionic EPS enriched with carboxyl-rich glycoproteins due to glucose induction and rice bran lignocellulosic components, thereby bridging for maximum turbidity removal [36]. *P. chrysosporium*, a lignin specialist, produced polysaccharide-heavy EPS primarily from lignocellulose degradation, supporting moderate performance through adsorption but limited bridging [37]. Strain *P. chrysosporium* excels at capturing individual particles (adsorption) but lacks the long-chain glycoprotein network required for inter-particle bridging, as *Trichoderma* sp. provides. Strain *C. utilis*, as a yeast relying on simple carbon metabolism, yielded mannan-dominant EPS [33], which favored adhesion over effective flocculation, resulting in lower removal efficiency. Thus, the lignocellulosic substrate (rice bran), combined with glucose and malt extract, likely enriched Trichoderma EPS with carboxyl-rich glycoproteins, thereby enhancing its coagulation superiority over the others.

Overall, turbidity reduction differed significantly among the three microbes on the measured days (Substrate 2), as indicated by a one-way ANOVA ($p=0.026 < 0.05$). Strain *Trichoderma* sp. exhibits the strongest turbidity removal (mean 62.8), evidencing its ideal ability to produce a bio-coagulant that can be used for turbid water. Strain *C. utilis* lags notably (mean 51.0), possibly due to slower metabolic rates, while *P. Chrysosporium* remains intermediate and stable (std 4.2). Tukey HSD reveals a significant difference in turbidity removal between *Trichoderma* sp. and *C. utilis* (mean diff -11.8, $p=0.030$), but not between *Trichoderma* sp. and *P. chrysosporium* ($p=0.073$) or *C. utilis* and *P. chrysosporium* ($p=0.872$).

An independent *t*-test was also conducted to compare turbidity removal efficiency between two microbial substrates (Substrate 1 mean 20.6 vs. Substrate 2 mean 55.6), revealing a statistically significant difference ($t = 10.60$, $p < 0.001$). These results confirm that the growth

conditions in Substrate 2 substantially enhanced the turbidity removal performance of the three microbes.

Our findings thus provide the first evidence that these *Trichoderma* sp. and *C. utilis* can achieve 71% and 59% turbidity removal, respectively, expanding the portfolio of microbial candidates for eco-friendly coagulation.

3.4. Scanning Electron Microscopy (SEM)

A morphological structure of microorganisms formed on rice bran was observed using scanning electron microscopy. The SEM micrographs of microorganisms are presented in Figure 6. After growth on the rice bran substrate for seven days at 30ºC, the microorganisms colonized the substrate surface.

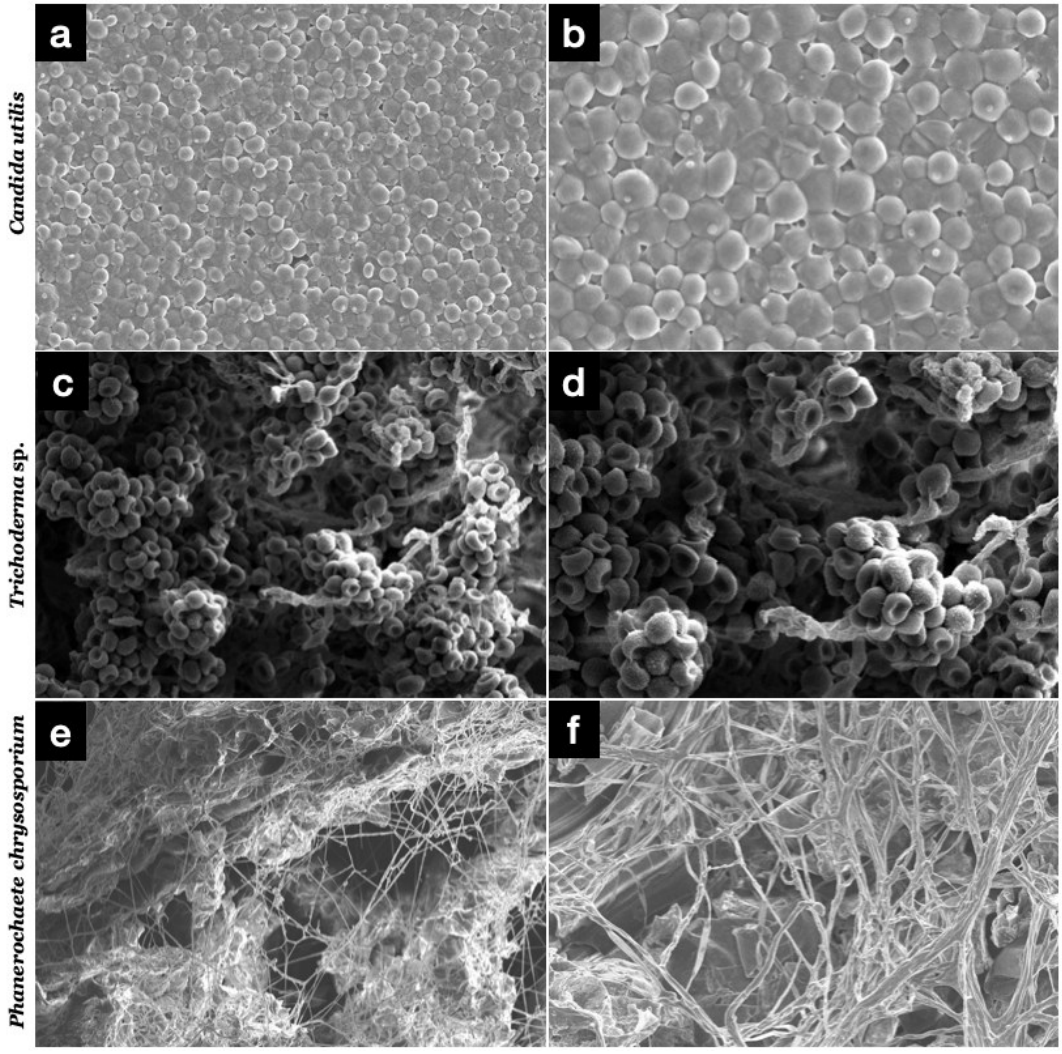


Figure 6. Scanning Electron Micrographs of Microbe Colonies on Rice Bran (a) *Candida utilis* [(a)×1000, (b)×2000], *Trichoderma* sp. [(c)×2000, (d)×3000], and *Phanerochaete chrysosporium* [(e)×180, (f)×450].

SEM examination showed that the rice bran substrate was completely colonized by *Candida utilis* yeast cells (Figure 6(a)). The shape of *Candida utilis* cells varies from circular to ellipsoidal/oval, and their length ranges from 3 to 6 µm (Figure 6(b)). Furthermore, some polar budding was observed, as shown in Figure 6(b).

Figure 6 (c) and 6 (d) show conidial balls developed from terminal phialides. The *Trichoderma* sp. conidial surface morphology appeared smooth when observed in SEM at a magnification of ×2000 (Figure 6 (c)) and in optical microscopy (Figure 1e, 1f), whereas it was more distinctly roughened and verrucose when observed in SEM at a magnification of ×3000 (Figure 6 (d)).

SEM also displayed *P. chrysosporium* growth on the surface of the rice bran substrate. The fungal mycelium of *P. chrysosporium* was filamentous and interconnected to form a spatial network. Rice bran substrate was colonized and penetrated by *P. chrysosporium*, as presented in Figure 6 (e). The hyphae of *P. chrysosporium* may be seen clearly in Figure 6 (f). Also, some spores were seen at the hyphae's terminal.

4. CONCLUSION

By introducing *Trichoderma sp*. and *C. utilis* as new fungal biocoagulants, this work opens a promising avenue for developing sustainable, microbe-based coagulation strategies in water treatment. Solid-state fermentation of *Trichoderma* sp. on the rice bran substrate proved particularly promising for producing effective natural coagulants. The results also showed that *Trichoderma* sp. was more effective at removing turbidity than the other coagulants (*Candida utilis* and *Phanerochaete chrysosporium*). Furthermore, solid-state fermentation with low water content is more advantageous than that with high water content for producing greater numbers of microorganisms, thereby yielding a high concentration of coagulants. SEM micrographs indicate that *Candida utilis*, *Trichoderma* sp., and *Phanerochaete chrysosporium* effectively colonized the solid substrate. In this study, biocoagulants, rice bran (as an abundant raw material), and tray-based solid-state fermentation (SSF) were employed, collectively reducing operational costs and dependence on expensive and complex extraction procedures.

REFERENCES

[1] Okaiyeto K, Nwodo UU, Mabinya LV., Okoli AS, Okoh AI. (2016) Evaluation of flocculating performance of a thermostable bioflocculant produced by marine Bacillus sp. Environmental Technology (United Kingdom). 37(14):1829–42. https://doi.org/10.1080/09593330.2015.1133717

[2] Ang TH, Kiatkittipong K, Kiatkittipong W, Chua SC, Lim JW, Show PL, et al. (2020) Insight on Extraction and Characterisation of Biopolymers as the Green Coagulants for Microalgae Harvesting. Water, Mdpi. 12(1288):1–31. https://doi.org/10.3390/w12051388

[3] Ramavandi B. (2014) Treatment of water turbidity and bacteria by using a coagulant extracted from Plantago ovata. Water Resources and Industry. 6:36–50. https://doi.org/10.1016/j.wri.2014.07.001

[4] Jebun N, Al Mamun AA, Alam MZ, Raus RA, Nedjai R. (2025) Performance of Myco-coagulant from Lentinus squarrosulus for Turbidity Reduction in a Settling Column. Malaysian Journal of Fundamental and Applied Sciences. 21(1):1506–12. https://doi.org/10.11113/mjfas.v21n1.2743

[5] Nie Y, Wang Z, Zhang R, Ma J, Zhang H, Li S, et al. (2021) Aspergillus oryzae, a novel eco-friendly fungal bioflocculant for turbid drinking water treatment. Separation and Purification Technology. 279(September):119669. https://doi.org/10.1016/j.seppur.2021.119669

[6] Khalil H, Legin E, Kurek B, Perre P, Taidi B. (2021) Morphological growth pattern of Phanerochaete chrysosporium cultivated on different Miscanthus x giganteus biomass fractions. BMC Microbiology. 21(1):1–18. https://doi.org/10.1186/s12866-021-02350-8

[7] Lam WC, Pleissner D, Lin CSK. (2013) Production of fungal glucoamylase for glucose production from food waste. Biomolecules. 3(3):651–61. https://doi.org/10.3390/biom3030651

[8] Saykhedkar SS, Singhal RS. (2004) Solid-state fermentation for production of griseofulvin on rice bran using Penicillium griseofulvum. Biotechnology Progress. 20(4):1280–4. Doi:10.1021/bp0343662

[9] Chen X, Yan J, Chen J, Gui R, Wu Y, Li N. (2021) Potato pomace: An efficient resource for Monascus pigments production through solid-state fermentation. Journal of Bioscience and Bioengineering. 132(2):167–73. https://doi.org/10.1016/j.jbiosc.2021.03.007

[10] Yoon LW, Ang TN, Ngoh GC, Chua ASM. (2014) Fungal solid-state fermentation and various methods of enhancement in cellulase production. Biomass Bioenergy. 67:319–38. https://doi.org/10.1016/j.biombioe.2014.05.013

[11] Try S, Voilley A, Chunhieng T, De-Coninck J, Waché Y. (2018) Aroma compounds production by solid state fermentation, importance of in situ gas-phase recovery systems. Applied Microbiology and Biotechnology. 102(17):7239–55. https://doi.org/10.1007/s00253-018-9157-4

[12] Vandenberghe LPS, Karp SG, de Oliveira PZ, de Carvalho JC, Rodrigues C, Soccol CR. (2018) Solid-State Fermentation for the Production of Organic Acids. Current Developments in Biotechnology and Bioengineering. 415–34. DOI:10.1016/B978-0-444-63990-5.00018-9

[13] Banat IM, Carboué Q, Saucedo-Castañeda G, de Jesús Cázares-Marinero J. (2021) Biosurfactants: The green generation of speciality chemicals and potential production using Solid-State fermentation (SSF) technology. Bioresource Technology. 320:124222. https://doi.org/10.1016/j.biortech.2020.124222

[14] Wang F, Terry N, Xu L, Zhao L, Ding Z, Ma H. (2019) Fungal laccase production from lignocellulosic agricultural wastes by solid-state fermentation: A review. Microorganisms. 7(12). https://doi.org/10.3390/microorganisms7120665

[15] Luthfi WM, Alias H, Tay GS, Lee CK. (2020). Production and characterization of bioflocculant via solid state fermentation process using oil palm empty fruit bunch fibers as substrate. Biocatalysis and Agricultural Biotechnology, 23, 101454. https://doi.org/10.1016/j.bcab.2019.101454

[16] Al Mamun A, Nedjai R, Alam MZ. (2023) Scaling up the production of myco-coagulant using solid-state fermentation for water treatment. Malaysian Journal of Microbiology. 19(6):777–785. http://dx.doi.org/10.21161/mjm.230025

[17] Nedjai R, Al-Mamun A, Alam MZ. (2024) Effects of initial turbidity and myco-coagulant dose on the effectiveness of the coagulation process in water treatment. Applied Chemical Engineering. 7(2). https://doi.org/10.59429/ace.v7i2.1546

[18] Nedjai R, Al Mamun A, Alam MZ. (2025) Assessment of powder myco-coagulant extracted from Phanerochaete concrescens for water treatment. ASEAN Engineering Journal. 15(1):57–63. https://doi.org/10.11113/aej.v15.21469

[19] Sala A, Barrena R, Artola A, Sánchez A. (2019) Current developments in the production of fungal biological control agents by solid-state fermentation using organic solid waste. Critical Reviews in Environmental Science and Technology, 49(8):655-694. doi:10.1080/10643389.2018.1557497

[20] Janarny G, Gunathilake KDPP. (2020) Changes in rice bran bioactives, their bioactivity, bioaccessibility and bioavailability with solid-state fermentation by Rhizopus oryzae. Biocatalysis and agricultural biotechnology. 23:101510. https://doi.org/10.1016/j.bcab.2020.101510

[21] Yadav U, Choudhury PP. (2014) Biodegradation of sulfosulphuron in agricultural soil by Trichoderma sp. Letters in Applied Microbiology. 59(5):479–86. https://doi.org/10.1111/lam.12306

[22] Sumon AK., & Ju LK. (2025). Factors affecting production of Scopulariopsis brevicaulis spores for use in self-healing concrete. Bioprocess and Biosystems Engineering, 48(9):1481-1494. https://doi.org/10.1007/s00449-025-03189-3

[23] Hattori R, Matsumura A, Yamawaki K, Tarui A, Daimon H. (2014). Excess soil water impact on colonization and extraradical mycelium biomass production of arbuscular mycorrhizal fungi in soybean field. International Journal of Agricultural Policy and Research, 41-48.

[24] Nussbaumer M, Karl T, Benz JP. (2024). Quantification of fungal biomass in mycelium composites made from diverse biogenic side streams. Fungal Biology and Biotechnology, 11(1):20. https://doi.org/10.1186/s40694-024-00189-y

[25] Dharsana M, Prakash AJ. (2023). Nano-banana peel bio-coagulant in applications for the treatment of turbid and river water. Desalination and Water Treatment, 294:100-110. https://doi.org/10.5004/dwt.2023.29491

[26] Gebreabe ST, Worku A, Gnaro MA, Kassahun E, Ale TH, Sime T, et al. (2025). Response Surface Optimization of Turbidity Removal from Highly Turbid Surface Water Using Orange Peel Powder as a Sustainable Coagulant. Next Research, 101058. https://doi.org/10.1016/j.nexres.2025.101058

[27] Pu S, Ma H, Deng D, Xue S, Zhu R, Zhou Y, et al. (2018). Isolation, identification, and characterization of an Aspergillus niger bioflocculant-producing strain using potato starch wastewater as nutrilite and its application. PloS one, 13(1). https://doi.org/10.1371/journal.pone.0190236

[28] Te Biesebeke R, Ruijter G, Rahardjo YSP, Hoogschagen MJ, Heerikhuisen M, Levin A, et al. (2002) Aspergillus oryzae in solid-state and submerged fermentations: Progress report on a multi-disciplinary project. FEMS Yeast Research. 2(2):245–8. doi:10.1111/j.1567-1364.2002.tb00089.x

[29] November EJ, Van Impe JF. (2000) Evaluation of online viable biomass measurements during fermentations of Candida utilis. Bioprocess Engineering. 23(5):473–7. https://doi.org/10.1007/s004499900179

[30] Zheng R, Pan F. (2019) Online tendency control of dissolved oxygen concentration during aerobic fed-batch fermentations. Applied Sciences (Switzerland). 9(23). https://doi.org/10.3390/app9235232

[31] Mendes CE, Badino AC. (2015) Oxygen transfer in different pneumatic bioreactors containing viscous Newtonian fluids. Chemical Engineering Research and Design. 94(August):456–65. https://doi.org/10.1016/j.cherd.2014.09.002

[32] Abu Tawila ZM, Ismail S, Dadrasnia A, Usman MM. (2018). Production and characterization of a bioflocculant produced by Bacillus salmalaya 139SI-7 and its applications in wastewater treatment. Molecules, 23(10):2689. https://doi.org/10.3390/molecules23102689

[33] Gientka I, Bzducha-Wróbel A, Stasiak-Różańska L, Bednarska AA, Błażejak S. (2016). The exopolysaccharides biosynthesis by Candida yeast depends on carbon sources. Electronic Journal of Biotechnology, 22, 31-37. https://doi.org/10.1016/j.ejbt.2016.02.008

[34] Li N, Liu J, Yang R, Wu L. (2020). Distribution, characteristics of extracellular polymeric substances of Phanerochaete chrysosporium under lead ion stress and the influence on Pb removal. Scientific Reports, 10(1):17633. https://doi.org/10.1038/s41598-020-74983-0

[35] Li H, Yu H, Zhu H. (2017). Structure studies of the extracellular polysaccharide from Trichoderma sp. KK19L1 and its antitumor effect via cell cycle arrest and apoptosis. Applied biochemistry and biotechnology, 182(1):128-141. https://doi.org/10.1007/s12010-016-2315-1

[36] Rubeena M, Neethu K, Sajith S, Sreedevi S, Priji P, Unni KN, et al. (2013). Lignocellulolytic activities of a novel strain of Trichoderma harzianum. Advances in Bioscience and Biotechnology, 4(02):214-221. https://doi.org/10.4236/abb.2013.42030

[37] Khalil H, Legin E, Kurek B, Perre P, Taidi B. (2021). Morphological growth pattern of Phanerochaete chrysosporium cultivated on different Miscanthus x giganteus biomass fractions. BMC microbiology, 21(1):318. https://doi.org/10.1186/s12866-021-02350-8

https://doi.org/10.31436/iiumej.v27i2.4237

Optimization and Scale-Up of Solid-State Bioconversion for Microbial Coagulant Production toward Sustainable Water Treatment

MAROUA FELLAH[1*], MD ZAHANGIR ALAM[1], ABDULLAH AL MAMUN[2], ABDELMOHSEN BENOUDJIT[3], NASSERELDEEN AHMED KABBASHI[1], NURUL SAKINAH ENGLIMAN[1]

[1]*Department of Biotechnology Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, Jalan Gombak, Kuala Lumpur, 53100, Selangor, Malaysia*
[2]*Department of Civil Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, Jalan Gombak, Kuala Lumpur, 53100, Selangor, Malaysia*
[3]*Department of Ecology and Environment, Kulliyyah of Natural Science and Life University (Université Batna 2), 53, Route de Constantine. Fsdis, Batna, 05000, Batna Province, Algeria*

**Corresponding author: fellahmaroua06@gmail.com*

(Received: 23 February 2026; Accepted: 21 April 2026; Published online: 10 May 2026)

ABSTRACT: Microbial coagulants are emerging as an eco-friendly alternative to chemical coagulants owing to their non-toxicity, biodegradability, and sustainability. However, their use is limited by high production costs linked to expensive nutrient media and scalability issues. To address these challenges, this study developed a cost-effective microbial coagulant based on a fungus using cocopeat as a substrate. Growth conditions (malt extract, glucose, and pH) were optimized using the Face-Centered Central Composite Design (FCCCD) under Response Surface Methodology (RSM) to maximize coagulant yield. The performance in water treatment was evaluated using jar tests, and large-scale production techniques were investigated using a tray bioreactor and an agitated drum bioreactor (ADB). The fungal strain used in this study was identified as *Phanerochaete concrescens*, which demonstrated high turbidity-reducing efficiency, achieving a maximum flocculating activity of 93.06% in a standard jar test. These results were obtained under statistically optimized growth conditions, specifically 5 days of incubation at 25°C, 70% moisture content, 3% malt extract, 2.5% glucose, and pH 7, confirming model predictions. *Phanerochaete concrescens* was successfully scaled using a tray fermenter, achieving a flocculation rate of 92% after 5 days of incubation at 70% moisture, pH 7, and a 3 cm substrate thickness. However, in a 30 L agitated bioreactor drum, flocculating activity was recorded at 29% due to insufficient fungal growth at the tested agitation rate. This study confirms the feasibility of a cost-effective microbial coagulant that can be scaled appropriately, offering a sustainable and biodegradable alternative to chemical coagulants.

ABSTRAK: Koagulan mikrob muncul sebagai alternatif mesra alam pada koagulan kimia kerana sifatnya yang tidak toksik, terbiodegradasi, dan mampan. Walau bagaimanapun, penggunaannya terhad kerana kos pengeluaran tinggi disebabkan oleh media nutrien yang mahal dan isu kebolehskalaan. Bagi menangani cabaran ini, kajian ini membangunkan koagulan mikrob kos efektif berdasarkan kulapuk kokopit sebagai substrat. Keadaan pertumbuhan (ekstrak malt, glukosa, dan pH) dioptimumkan menggunakan Reka Bentuk Komposit Berpusat Muka (FCCCD) di bawah Kaedah Gerak Balas Permukaan (RSM) bagi memaksimum hasil koagulan. Prestasi dalam rawatan air dinilai menggunakan ujian balang, dan teknik pengeluaran skala besar dikaji menggunakan bioreaktor dulang dan bioreaktor dram bergoncang (ADB). Strain kulat yang digunakan dalam kajian ini dikenal pasti sebagai Phanerochaete concrescens, menunjukkan kecekapan tinggi dalam mengurangkan kekeruhan,

https://doi.org/10.31436/iiumej.v27i2.4237

mencapai aktiviti pengflokulan maksimum 93.06% melalui ujian piawai balang. Keputusan ini diperoleh di bawah keadaan pertumbuhan yang optimum secara statistik, iaitu 5 hari pengeraman pada suhu 25°C, kandungan lembapan 70%, ekstrak malt 3%, glukosa 2.5%, dan pH 7, mengesahkan model ramalan. Phanerochaete concrescens berjaya diskala menggunakan penapai dulang, dengan kadar pengflokulan 92% dicapai selepas 5 hari pengeraman di bawah lembapan 70%, pH 7, dan ketebalan substrat 3 cm. Walau bagaimanapun, dalam dram bioreaktor bergoncang 30 L, aktiviti pengflokulan direkodkan pada 29% disebabkan oleh pertumbuhan kulat yang tidak mencukupi pada kadar goncangan yang diuji. Kajian ini mengesahkan kebolehlaksanaan koagulan mikrob kos efektif pada skala wajar, menawarkan alternatif mampan dan terbiodegradasi pada koagulan kimia.

KEYWORDS: *microbial coagulant, response surface methodology, solid-state bioconversion, scale-up, water treatment.*

1. INTRODUCTION

Water is vital for human survival and ecological balance, underscoring the global necessity of universal access to clean, safe water. However, population growth, industrialization, and water scarcity due to climate change pose substantial challenges to water resources, underscoring the need for innovative and sustainable water treatment technologies [1, 2]. Several technologies have been employed in water treatment, ranging from conventional methods such as sedimentation and filtration to advanced techniques such as reverse osmosis, ozonation, and membrane filtration [3, 4, 5, 6]. Among these, the coagulation-flocculation process is valued for its simplicity, cost-effectiveness, and efficiency in water treatment, removing organic matter, suspended solids, and heavy metals [7]. The coagulation-flocculation process involves adding coagulants to neutralize the surface charges of suspended particles, facilitating their aggregation into larger flocs that can be removed by sedimentation, thereby enhancing water clarity and quality [8,9,10]. Coagulants can be categorized into chemical coagulants (inorganic, such as aluminum sulfate and ferric chloride), synthetic organic coagulants (such as polyacrylamide derivatives), and natural coagulants (plant, animal, and microbial origin).

Water treatment has traditionally used chemical coagulants, both inorganic (such as polyaluminum chloride and ferric chloride) and organic (such as polyacrylamide). Despite their widespread use, they pose significant environmental and health risks due to low biodegradability and potential toxicity, as well as concerns regarding cost and local availability [11,12]. Consequently, there has been an increasing trend towards seeking safer, sustainable alternatives that can effectively demonstrate coagulation-flocculation performance while minimizing environmental impact and health risks [13]. Natural coagulants are organic, non-toxic substances sourced from plants (e.g., starch, guar gum, Moringa oleifera), animals (e.g., shellfish extracts, chitosan), and microorganisms (e.g., bacteria, yeast, fungi, algae). They are regarded as sustainable and can be locally available. Within this broad category, microbial coagulants are emerging as a promising sustainable alternative to chemical coagulants, largely due to their rapid reproducibility, biodegradability, environmental friendliness, and effectiveness across applications such as heavy metal removal, dye wastewater treatment, drinking water purification, and microalgae harvesting [14, 15]. Microbial coagulants produce a wide range of biomolecules, including proteins, glycoproteins, polysaccharides, and extracellular polymeric substances (EPS), which effectively remove suspended particles, organic matter, and contaminants [16, 17].

Most research on microbial coagulants is largely limited to laboratory studies, highlighting a considerable challenge for large-scale industrial production and application. The main obstacle hindering this advancement is the high production cost. To reduce production and application costs, various strategies have been implemented, including exploring low-cost alternative media. Solid-state bioconversion is identified as a cost-effective alternative that reduces overall resource demands by employing low-cost agro-industrial residues, enhancing growth conditions, and using affordable lignocellulosic substrates, relying on specific microorganisms to convert these solid substrates into high-value products [18,19]. Among microorganisms, filamentous fungi are well-suited for solid-state bioconversion due to their strong enzymatic activity and distinctive hyphal growth, which facilitate effective penetration and nutrient assimilation in solid matrices, thereby enhancing substrate conversion efficiency and yield [20, 21, 22].

In the present study, a filamentous fungus was used to produce an effective microbial coagulant via solid-state bioconversion of cocopeat, a natural, low-cost agricultural by-product, with the aim of optimizing growth conditions to maximize microbial coagulant yield and turbidity removal efficiency at the laboratory scale [23,24]. The optimization employed Face-Centered Central Composite Design under response surface methodology to assess the key variables influencing microbial coagulant production, including malt extract, glucose, and pH. Additionally, these optimized parameters were applied in the first reported scale-up of this process, utilizing several bioreactor configurations, including a tray bioreactor and an agitated bioreactor drum (ABD). This study demonstrates that using agricultural residues and low-nutrient cultivation methods can substantially lower production costs without compromising effectiveness. Furthermore, this sustainable and eco-friendly approach shows strong potential for industrial-scale up, thereby enhancing the commercial feasibility of microbial coagulant production for water treatment.

2. Materials and Methods

2.1. Microorganism collection

The present study employed a fungus as a microorganism that was obtained from the Biotechnology Laboratory of the International Islamic University of Malaysia (IIUM). The strain was preserved through successive subcultures on potato dextrose agar (PDA) and maintained on agar at 4°C.

2.2 Molecular identification of microbial coagulant-producing fungal strain

DNA samples were extracted from a one-week-old PDA pure culture of a fungal strain for molecular identification. The extraction, PCR amplification of ITS, purification, and sequencing were carried out by Apical Scientific MBS Laboratories in Malaysia. Genomic DNA was sequenced using the 18S rRNA gene, ITS1, 5.8S rRNA gene, ITS2, and 28S rDNA gene. These sequences were then compared to related sequences using BLAST in GenBank, leading to the construction of a phylogenetic tree via pairwise alignments and the neighbor-joining method.

2.3. Preparation of inoculum (mycelial suspension)

In this study, the inoculum was prepared by harvesting fungal cultures after 7 days of incubation, washing them with 20 ml of sterile distilled water to create a uniform mycelial suspension for solid-state bioconversion, ensuring consistency in coagulant production.

https://doi.org/10.31436/iiumej.v27i2.4237

2.4. Substrate collection and preparation

In this study, cocopeat, sourced from the International Islamic University Malaysia, was processed into a fine powder, sieved, and stored for fermentation. This method utilizes a cost-effective agricultural byproduct, enhancing sustainability and reducing material costs.

2.5. Experimental design

In this study, the production of the desired microbial coagulant was systematically optimized. Growth conditions were designed and analyzed using a Face-Centered Central Composite Design (FCCCD) within Response Surface Methodology (RSM) via Design Expert® 7.0.0. The total number of samples examined under RSM was 20 experimental runs, each analyzed in triplicate, including six center points, as shown in Table 1.

Table 1. Response surface design (FCCCD) of three independent parameters

Run	Malt extract [% w/w]	pH	Glucose [% w/w]
1	2.00	4.00	1.00
2	4.00	4.00	1.00
3	2.00	8.00	1.00
4	4.00	8.00	1.00
5	2.00	4.00	3.00
6	4.00	4.00	3.00
7	2.00	8.00	3.00
8	4.00	8.00	3.00
9	2.00	6.00	2.00
10	4.00	6.00	2.00
11	3.00	4.00	2.00
12	3.00	8.00	2.00
13	3.00	6.00	1.00
14	3.00	6.00	3.00
15	3.00	6.00	2.00
16	3.00	6.00	2.00
17	3.00	6.00	2.00
18	3.00	6.00	2.00
19	3.00	6.00	2.00
20	3.00	6.00	2.00

The flocculating activity rate (%) was selected as the design experiment's response (Y). For the variable assessments, the independent factors were adjusted at three levels: low (−1), medium (0), and high (+1). The parameters employed were malt extract concentration (2-4 % w/w), pH (4-8), and glucose concentration (1-3 % w/w). The relationship between the flocculation activity rate and independent factors was modeled using a second-order polynomial equation:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 + \beta_{11} {X_1}^2 + \beta_{22} {X_2}^2 + \beta_{33} {X_3}^2 \quad (1)$$

where Y indicates the flocculation activity rate, while X_1, X_2, and X_3 stand for malt extract, pH, and glucose, respectively. The coefficients include β_0 (intercept), β_1, β_2, and β_3 (linear coefficients), β_{12}, β_{13}, and β_{23} (interaction coefficients), and β_{11}, β_{22}, and β_{33} (quadratic coefficients). The optimal conditions were determined using analysis of variance (ANOVA) and contour plots, with model adequacy evaluated by the coefficient of determination (R^2) to assess goodness of fit [25]. For each experimental run, a microbial coagulant was produced via

solid-state bioconversion by moistening 3 g of cocopeat with 8 mL of a variable-composition production medium. This medium's malt extract, glucose concentrations, and pH were adjusted for each run as per the experimental design matrix. The medium was sterilized at 120°C for 15 minutes, inoculated, and then incubated at 25°C for five days.

2.6. Extraction process of microbial coagulant

In this study, a modified method from [26] was used to extract the produced microbial coagulant after the incubation period. The procedure for obtaining the crude extract involved adding 2 g of the fermented substrate and 20 ml of pH 7.0 buffer, shaking at 250 rpm for 1 hour at 30°C, and then filtering.

2.7. Jar procedure for microbial coagulant performance assessment

The jar test method using an SW6 UK Flocculator was used to assess the effectiveness of the produced microbial coagulant in reducing turbidity. This procedure involved adding 10 mL of the extracted microbial coagulant to a 300 mL sample. The sample was rapidly mixed for 7 minutes at 250 rpm. This was followed by a slow mixing process for 22 minutes at 90 rpm. The mixture was then left to settle for 1 hour. The turbidity of the treated water and raw water samples was determined using a turbidity meter, Model 2100Q. The rate of flocculating activity was determined from kaolin suspension tests using the method proposed by [27], where A is the initial turbidity (NTU) and B is the final turbidity (NTU). All tests were performed in triplicate.

$$Flocculating\ activity\ rate\ (\%) = \frac{A-B}{A} \times 100 \quad (2)$$

2.8. Experimental validation

Design-Expert software was used to optimize experiments based on the parameter values. Manual adjustments were made based on the equations derived from the RSM simulation to maximize the flocculating activity rate (%), as shown in Table 2. A comparison of predicted values and experimental results demonstrated the model's accuracy in predicting flocculation activity rates (%) using percentage error analysis.

Table 1 . Validation for the FCCCD model

Run	Malt extract [% w/w]	pH	Glucose [% w/w]
1	4	8	3
2	3	7	2.5
3	3	7.5	2

2.9. Scale-up procedure of microbial coagulant

For the production of microbial coagulant, two types of bioreactors, namely a tray bioreactor (Figure 1(a) and a 30 L agitated drum bioreactor (Figure 1(b), have been used to carry out the scale-up process for large-scale production through solid-state bioconversion.

2.9.1. Microbial coagulant scale-up in Tray Bioreactor (TB)

The first scaled-up process for the target microbial coagulant involved using a tray bioreactor, where a sterilized cocopeat substrate mixture was added and incubated for 5 days at 25°C. Some of the critical factors that are part of the process of scaling up this coagulant, such as the moisture content (60-80%) of the substrate bed, the substrate bed thickness (1-3

cm), and the pH level of the medium (7-9), were also tested to assess their effect on the coagulant produced.

2.9.2. Microbial coagulant scale-up in Rotary Drum Bioreactor (RDB)

For the large-scale production of the desired microbial coagulant through solid-state bioconversion, the process was carried out using a 30 L rotary drum bioreactor with 10 kg of sterilized cocopeat substrate at 70% moisture level, and the stirring process was performed for 2 hours a day over 5 days using the optimized growth condition at laboratory scale. Daily measurement of coagulant efficiency was performed using a jar test.



Figure 1. Scale-up procedure of microbial coagulant, a) tray foil, b) rotatory drum bioreactor.

3. RESULTS AND DISCUSSIONS

3.1. Molecular identification of fungal strain

The potential fungal strain used in this study was identified as belonging to the genus *Phanerochaete concrescens.* Molecular identification based on the rRNA operon gene (18S rRNA, ITS1 spacer, 5.8S rRNA, ITS2 spacer, and 28S rRNA) sequences showed the highest similarity to *Phanerochaete concrescens*. Sequence similarity analysis of the fungal rRNA operons was conducted using 10 related strains of *Phanerochaete concrescens* from GenBank. The isolated filamentous strain exhibited 100% sequence identity with the type strain of *Phanerochaete concrescens*. Morphological and phylogenetic analyses confirmed the strain's identification as *Phanerochaete concrescens*. A phylogenetic tree was constructed to compare this strain with similar sequences available in GenBank (Fig 2). The tree was generated from rRNA operon gene sequences using BLAST pairwise alignments and the neighbor-joining method, resulting in an unrooted tree. However, to the best of our knowledge, this fungus has not been reported as a microbial coagulant-producer. These findings identify a new fungal strain capable of producing microbial coagulants, highlighting its potential for eco-friendly coagulation and flocculation applications.

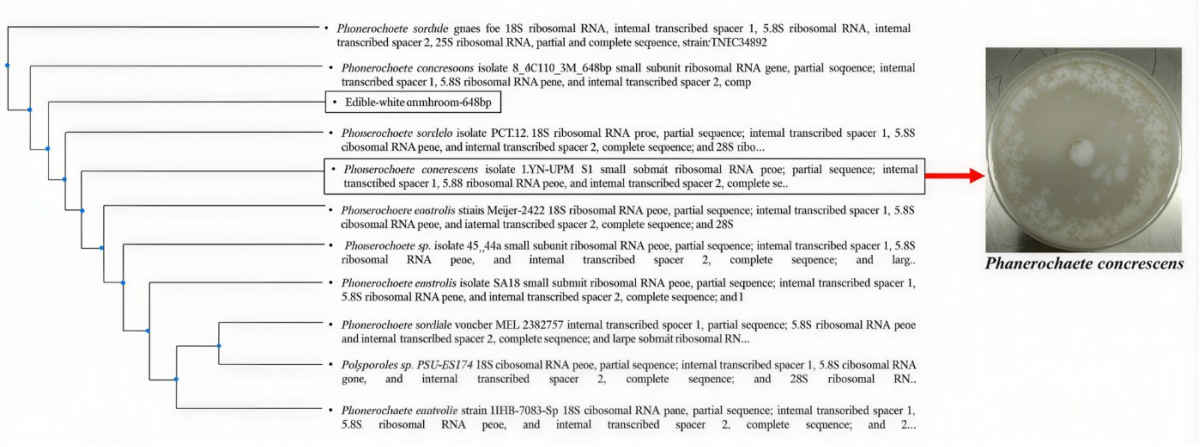


Figure 2. A phylogenetic tree based on 16 rRNA gene sequences of strain *Phanerochaete concrescens*.

3.2. Optimization of growth conditions for efficient microbial coagulant production

Solid-state fermentation using cocopeat as a lignocellulosic substrate is an effective method for producing a microbial coagulant and offers a sustainable alternative to traditional submerged fermentation [23, 24]. In the current research, response surface methodology (RSM) was employed to optimize microbial coagulant production and assess its efficacy in turbidity removal, thereby revealing its potential as a sustainable water treatment technology. The Face-Centered Central Composite Design (FCCCD) was employed in the RSM approach to effectively explore nonlinear variable relationships among the variables, thereby enabling the optimization of factors that enhanced microbial coagulant yield and flocculating activity. Three critical independent variables, including malt extract concentration, pH, and glucose concentration, were selected based on their significant influence on the growth and metabolite production of the fungal strain responsible for coagulant production [23, 24].

As shown in Table 3, the flocculating activity rate varied significantly across experimental runs, ranging from 79.00% (run 2) to 93.03% (run 8), indicating the sensitivity of the fungal strain's metabolism to both nutrient composition and environmental pH. The range also indicates that, although the fungal strain exhibited the lowest flocculating activity at 79% under unfavorable conditions, the experiment underscores the importance of precision in controlling environmental conditions. Conversely, the highest percentage of flocculating activity, at 93.03%, underscores the effectiveness of the synergism between the optimized nutrient composition and the pH achieved during the experiment. Additionally, the repetition of activity levels at the central points of the experiment validated its reliability.

The optimization of experimental conditions was performed to enhance the flocculation activity rate (%), improve statistical interactions between factors, and reduce experimental costs. A second-order polynomial model was used to describe the empirical relationship between the dependent variable (flocculation activity rate (%)) and the independent variables: malt extract (A), pH (B), and glucose (C). The resulting regression equation was

$$Y(Flocculation\ activity\ rate, \%) = +90.45 - 1.04A + 1.19B + 1.77C - 1.06A^2 - 0.93B^2 - 0.59C^2 + 1.92AB - 2.29AC - 0.97BC \quad (3)$$

https://doi.org/10.31436/iiumej.v27i2.4237

Table 2. FCCCD experimental design and the response

Run	Malt extract [% w/w]	pH	Glucose [% w/w]	Flocculation activity rate (%)	
				Experimental	Predicted
1	2.00	4.00	1.00	89.36	89.19
2	4.00	4.00	1.00	79.00	78.68
3	2.00	8.00	1.00	89.19	89.68
4	4.00	8.00	1.00	86.83	86.85
5	2.00	4.00	3.00	90.07	90.10
6	4.00	4.00	3.00	89.19	88.75
7	2.00	8.00	3.00	86.33	86.70
8	4.00	8.00	3.00	92.80	93.03
9	2.00	6.00	2.00	91.14	90.43
10	4.00	6.00	2.00	87.83	88.34
11	3.00	4.00	2.00	87.43	88.33
12	3.00	8.00	2.00	91.81	90.71
13	3.00	6.00	1.00	88.10	88.09
14	3.00	6.00	3.00	91.82	91.63
15	3.00	6.00	2.00	91.00	90.45
16	3.00	6.00	2.00	89.50	90.45
17	3.00	6.00	2.00	88.24	90.45
18	3.00	6.00	2.00	90.76	90.45
19	3.00	6.00	2.00	91.23	90.45
20	3.00	6.00	2.00	91.54	90.45

Table 3. ANOVA of the developed model

Source	Sum of Squares	Mean Square	*F*-Value	*p*-Value
Model	160.00	17.78	15.40	< 0.0001
A-malt extract	10.90	10.90	9.44	0.0118
B-pH	14.18	14.18	12.29	0.0057
C-glucose	31.44	31.44	27.23	0.0004
AB	29.45	29.45	25.51	0.0005
AC	41.91	41.91	36.30	0.0001
BC	7.59	7.59	6.57	0.0282
A^2	3.10	3.10	2.68	0.1324
B^2	2.36	2.36	2.04	0.1833
C^2	0.95	0.95	0.82	0.3867
Lack of Fit	3.59	0.72	0.45	0.7981

R^2 = 0.9327, **Adjusted R^2** = 0.8721, **CV** = 1.21, **Adequate precision** = 18.879

The analysis of variance (ANOVA) of the developed model, summarized in Table 4, was employed to assess the adequacy and statistical validity of the proposed regression model. The F-value and corresponding p-value are key indicators for evaluating both the overall model significance and the interactions among the independent variables. The model exhibited an F-value of 15.40 with a p-value < 0.0001, indicating that the probability of such a large F-value arising from random noise is less than 0.01%, thereby confirming the model's statistical significance. According to standard statistical criteria, model terms with p-values < 0.05 are considered significant, whereas values > 0.1 indicate insignificance. As shown in Table 4, the linear terms malt extract (A), pH (B), and glucose (C) exerted significant effects on flocculating activity ($p < 0.05$). In addition, the interaction terms AB, AC, and BC were statistically significant, indicating strong synergistic interactions among the studied factors. In contrast, the quadratic terms A^2, B^2, and C^2 were not significant ($p > 0.1$), suggesting that pronounced curvature effects were not present within the investigated experimental domain. The lack-of-

fit test was non-significant (p = 0.7981), indicating good agreement between the experimental and predicted responses and confirming the model's suitability. The model's efficiency and predictive accuracy were further supported by the high coefficient of determination (R^2 = 0.9327) and adjusted R^2 (0.8721), indicating that 93.27% and 87.21% of the total variation in flocculating activity were explained by the model, respectively. The adequate precision ratio, which measures the signal-to-noise ratio, was 18.879, well above the desirable threshold of 4, indicating an adequate and reliable signal. Moreover, the low coefficient of variation (CV = 1.21%) reflects high experimental precision and reproducibility. Commonly, a model is deemed reasonably reproducible 79if its coefficient of variation (CV) is less than 10% [28].

3.3. Analysis of the response surfaces model

To further elucidate the interactions among medium components and their concentrations in optimizing microbial coagulant production, response surface methodology was employed. The relationships between variables were visualized using three-dimensional (3D) response surface plots and two-dimensional (2D) contour plots. The elliptical contour patterns observed in these plots confirmed the presence of significant interactions among the tested parameters, consistent with the ANOVA results.

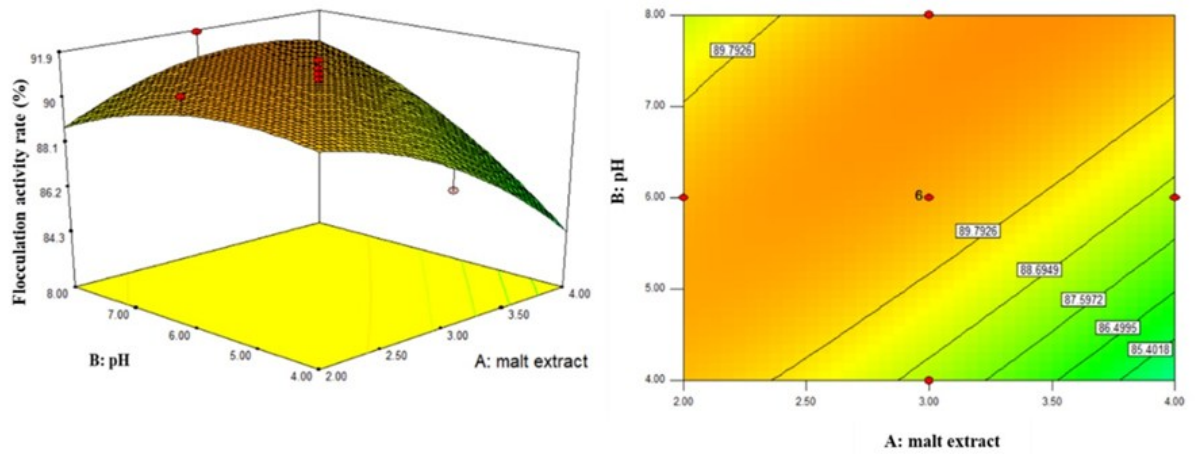


Figure 3. 3D and contour plots representing the effect of malt extract and pH on flocculation activity.

Figure 3 illustrates the interaction effect of pH and malt extract concentration on the flocculation activity rate (%), providing insight into the optimal conditions for achieving maximum flocculation efficiency. As shown in Figure 1, it is clear that the flocculation activity rate is significantly affected by both pH and malt extract concentration. The maximum flocculation activity of about 91.9% was achieved at a moderate pH range (6.0–7.5) and an optimal malt extract concentration of 2.5–3.5. The increase in pH was found to improve the flocculation activity, which can be attributed to enhanced charge neutralization and polymer bridging mechanisms. These results are in agreement with the study by [29], whereby *Aspergillus flavus* MCB 271 and *Aspergillus niger* MCBF 08 produced a bioflocculant optimally at pH 7. However, excessive malt extract concentrations reduced flocculation efficiency, which may be attributed to overdosing that hinders effective particle aggregation. The interaction between pH and malt extract concentration is crucial for determining the flocculation response, underscoring the importance of optimizing both factors simultaneously. Statistical analysis confirmed that the interaction between pH and malt extract significantly affected the flocculation rate, as indicated by the corresponding p-values presented in Table 4. These results emphasize the synergistic influence of pH and malt extract on flocculation

https://doi.org/10.31436/iiumej.v27i2.4237

performance and validate the response surface model as an effective tool for optimizing process conditions.

Figure 4 presents the interactive effect of malt extract and glucose on the flocculation activity rate. The highest response values, reaching approximately 90.6%, were observed within the orange region of the response surface, indicating the optimal concentration range for both factors. In contrast, further increases in either glucose or malt extract beyond this optimal range resulted in a noticeable decline in flocculation activity, as illustrated by the green region. These findings demonstrate that both malt extract and glucose play a crucial role in modulating the flocculation response, likely through their influence on microbial growth, metabolic activity, and extracellular polymer production. The existence of an optimal concentration range suggests that a balanced nutrient supply is essential for achieving maximum flocculation efficiency. At ideal concentrations, malt extract and glucose provide sufficient carbon and energy sources to support microbial metabolism without inducing substrate inhibition or metabolic imbalance. Conversely, excessive nutrient levels may disrupt microbial physiology or interfere with floc formation, thereby reducing flocculation performance. The findings were consistent with other studies, which reported that glucose is the most preferred carbon source [30]. However, other studies have shown that peptone is the preferred source of nitrogen [29].

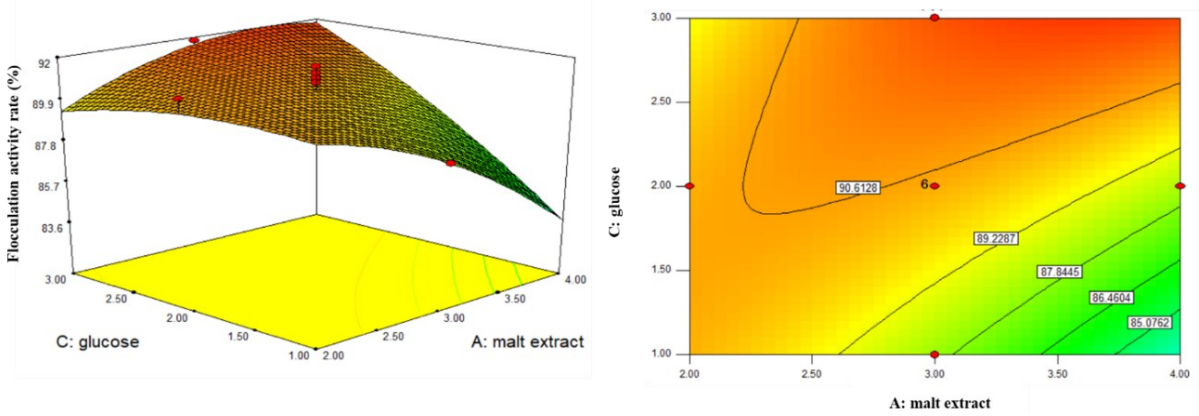


Figure 4. 3D and contour plots representing the effect of malt extract and glucose on flocculation activity.

Figure 5 shows the relationship between pH and glucose (m/m) concentration on the flocculation activity rate. The three-dimensional response surface displays a clear parabolic shape, confirming a quadratic relationship between these variables and the response. Both pH and glucose concentration positively influence flocculation activity, with the response increasing as these factors rise, up to an optimal point beyond which no further gains occur. The parabolic pattern between pH and glucose indicates that carbon flux through biosynthetic pathways depends heavily on enzyme stability within a narrow pH range, with slightly alkaline conditions promoting extracellular enzyme secretion and microbial coagulant formation [20]. This explains the decline in flocculation efficiency at extreme pH values, where enzymatic activity and metabolic balance may be disrupted. Additionally, the curvature of the response surface emphasizes the importance of maintaining optimal pH–glucose conditions to avoid substrate inhibition and metabolic stress. The close match between experimental data points and the predicted surface further confirms the good fit, accuracy, and predictive power of the developed model, demonstrating the effectiveness of response surface methodology for optimizing flocculation performance.

https://doi.org/10.31436/iiumej.v27i2.4237

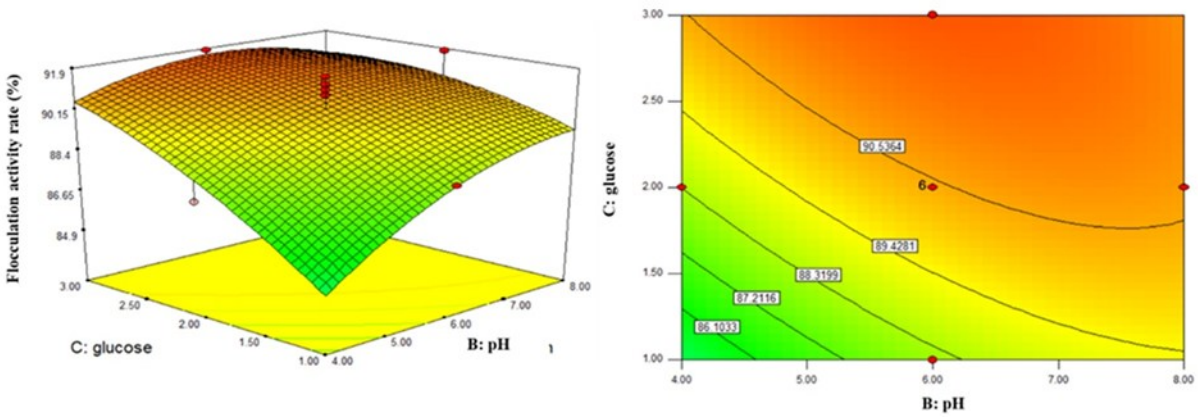


Figure 5. 3D and contour plots representing the effect of glucose and pH on flocculation activity.

3.4. Validation of the optimized parameter

Model validation evaluated the reliability of optimization outcomes by comparing predicted and experimental flocculating activity values. The Design Expert software's point prediction function assessed model accuracy, while high and low levels of key factors were manually adjusted to refine predictions of optimal conditions.

Additional extraction of the desired microbial coagulant was performed to validate the generated model and confirm the best results, as presented in Table 5. The optimized condition, obtained at 3% malt extract, 2.5% glucose, and pH 7, yielded a maximum flocculating activity of 93.06%, significantly reducing water turbidity from 672 to 46.63. However, in Run 3, a lower flocculating activity resulted in only a 60.55 NTU reduction. The experimental values were compared with the predicted values based on a standard error below 10%, confirming high predictive accuracy. The fact that these findings were so closely related proved that the response model accurately represented the predicted optimization.

Table 4. Validation of the experimental model

Run	Malt extract	pH	Glucose	Flocculation activity (%)		Error (%)
				Predicted	Experimental	
1	4	8	3	92.02	91.14	0.88
2	3	7	2.5	91.30	93.06	1.76
3	3	7.5	2	90.81	90.99	0.18

3.5. Large production of microbial coagulant

Significant progress has been made in researching microbial coagulants; however, their scaling up is still under development. In this context, the current study is conducted to enhance the production of microbial coagulant through solid-state bioconversion using different bioreactor types, including tray fermenters and agitated drum bioreactors, thereby providing an environmentally friendly solution for large-scale water treatment.

3.5.1. Large production of microbial coagulant in a tray fermenter

Tray fermenters are used economically for solid-state bioconversion, in which tray covers hold the substrates and inoculum for growth. The factors examined in this study included pH, moisture content, and substrate thickness, and a 5-day incubation at 25°C was used for coagulant-scale production by Phanerochaete concrescens.

3.5.1.1. Effect of pH value on flocculating activity rate (%) in a tray fermenter

One of the most influential factors is the medium's pH, which significantly affects the production of the microbial coagulant during solid-state bioconversion [31]. Thus, the effect of medium pH was evaluated over the range 6 to 9, as shown in Figure 6. Based on the obtained results, the highest flocculating activity was observed at 91.5% at pH 7 after 5 days of incubation. The flocculating activity decreased at acidic pH, while the moisture content (70 %) and agitation (3 cm) were considered. The optimal pH for bioflocculant production by A. niger was 7, as reported by [27, 32].

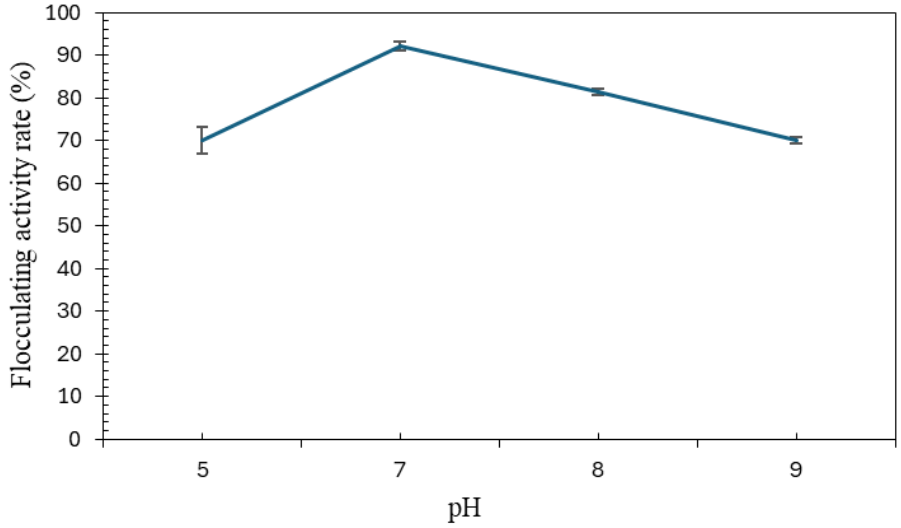


Figure 6. Effect of pH on flocculating activity rate (%) in a tray fermenter.

3.5.1.2. Effect of moisture content on flocculating activity rate (%) in a tray fermenter

For microbial coagulant production, moisture content is a crucial factor [33]. Thus, different moisture contents in the fermentation media were evaluated to assess the optimal level that supports better growth of the used fungal strain and, hence, better substrate utilization, resulting in maximum microbial coagulant production.

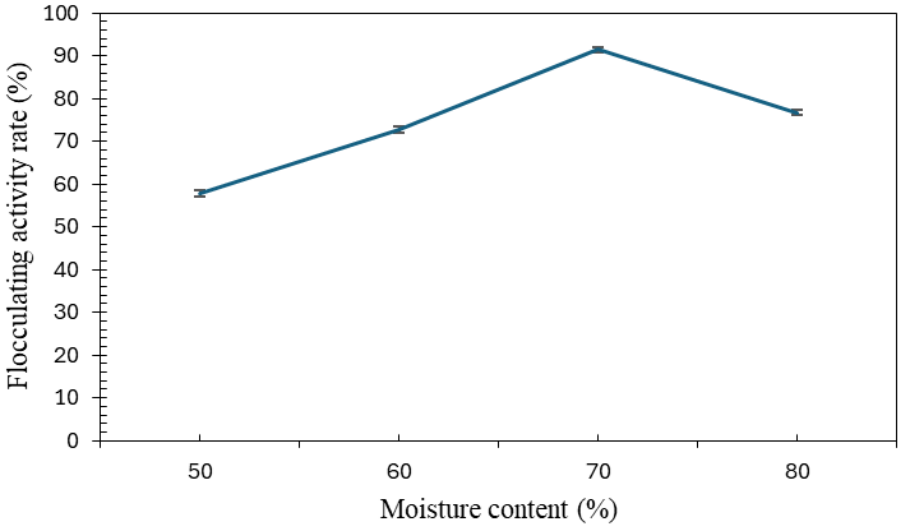


Figure 7. Effect of moisture content on flocculating activity rate (%) in the tray fermenter.

As indicated by the result presented in Figure 7, the highest flocculating activity of 91.5% was found at a moisture content of 70%, whereas the lowest flocculating activity rate of 54.5%

was noted at a moisture content of 50%. It was justified that *Phanerochaete concrescens*. required a considerable amount of medium moisture. This could be explained by the fact that the coco-peat substrate holds water. The moisture content facilitates proper substrate swelling, thereby providing microorganisms with an opportunity to use it effectively.

3.5.1.3. Effect of thickness of substrate on flocculating activity rate (%) in a tray fermenter

In this study, the thickness of the substrate layer was examined to assess its impact on the production of the target microbial coagulant during the solid-state bioconversion process. A layer thickness of 3 cm was found optimal, yielding a maximum microbial production of 92%. A layer thickness of 3 cm was found to be optimal for maximum microbial coagulant output. Substrate thicknesses of 1 and 2 cm resulted in lower flocculating activity rates of 60% and 71.2%, respectively (Figure 8). This could be due to lower thickness not supporting the full growth of fungal mycelia, resulting in reduced production yield. Conversely, increasing the layer thickness can hinder microbial respiration and aeration, leading to poor growth. Additionally, a thicker substrate can reduce product diffusion, potentially lowering enzyme activity due to feedback inhibition [34].

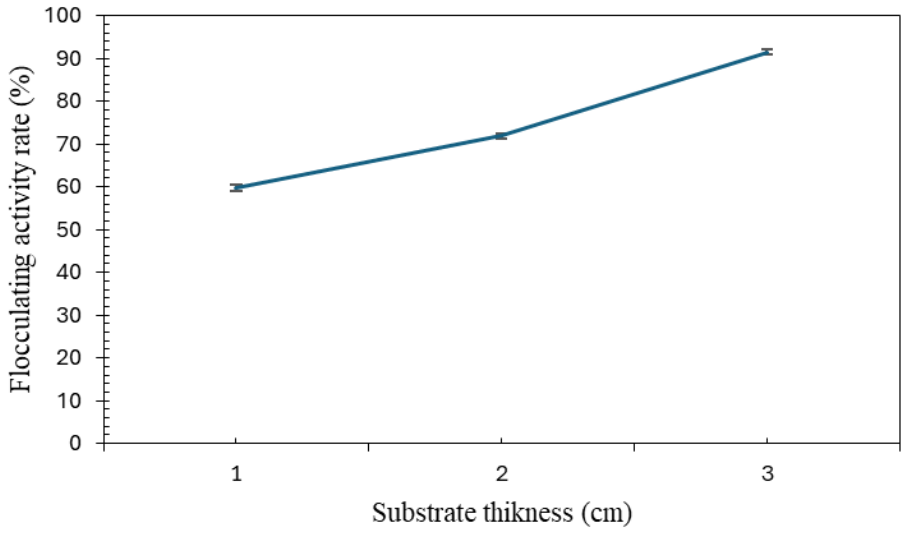


Figure 8. Effect of substrate thickness on flocculating activity rate (%) in the tray fermenter.

3.5.2. Large-scale production of microbial coagulant in an agitated drum bioreactor

Further research is underway to improve the design of the solid-state bioconversion bioreactor for producing the microbial coagulant, which is effective at removing water turbidity. However, scaling this experiment from the laboratory to industrial scale has been a significant challenge, with no prior attempt to increase production of this particular microbial coagulant. Therefore, this study is considered the first attempt to scale up the production of a microbial laccase using a 30 L agitated drum bioreactor. Scale-up production in a 30 L agitated drum bioreactor did not yield the desired microbial coagulant. Based on the results in Figure 9, flocculating activity ranged from 22 to 29 % over 7 days of incubation. It can be observed that within the first three days of incubation, the fungus grows in the substrate with daily mixing for 1 hour. However, no growth was observed on the fourth day of incubation. Therefore, it is clear that the stirring process negatively affected the growth of the fungus as the hyphae were damaged.

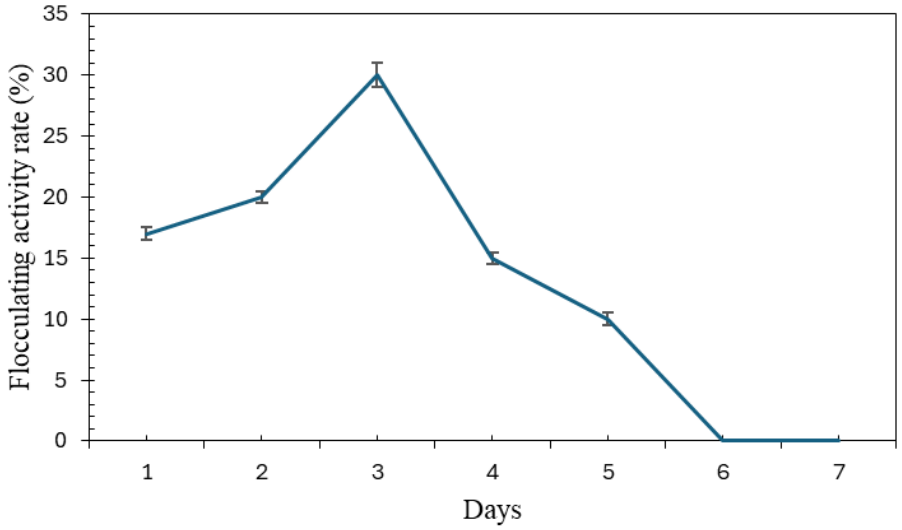


Figure 9. Flocculating activity rate (%) in an agitated drum bioreactor during 7 days.

4. CONCLUSIONS

In conclusion, the study presents a cost-effective method for producing microbial coagulants from cocopeat through solid-state bioconversion. Optimal conditions for large-scale production were established as 3% malt extract, 2.5% glucose, pH 7, 70% moisture, and 25°C for 5 days, resulting in 93.06% flocculating activity. A comparison of production methods revealed that the tray fermenter maintained 92% activity under these conditions, while the agitated drum bioreactor showed only 29% effectiveness. This approach uses agro-industrial waste and optimized fermentation to produce microbial coagulants for water treatment, providing a sustainable and cost-effective approach.

ACKNOWLEDGEMENT

The authors acknowledge the Islamic Development Bank (ISDB) Grant No. SPI20-001-0001 for funding this research and express gratitude to the International Islamic University Malaysia (IIUM), particularly the Research Management Center (RMC), for financial administration and project monitoring, along with the Department of Biotechnology Engineering, IIUM, for providing laboratory facilities.

REFERENCES

[1] Gizewski J., van der Sande L., Holtmann D. (2023) Contribution of electrobiotechnology to sustainable development goals. Trends Biotechnol. 41(9):1106–1108. https: //doi.org/10.1016/j.tibtech.2023.02.009

[2] Qu J., Liu H., Liu G. (2022) Science and technology for combating global water challenges. Engineering 9:1–2. https://doi.org/10.1016/j.eng.2022.01.007

[3] Chakraborty R., Asthana A., Singh A.K., Jain B., Susan A.B.H. (2022) Adsorption of heavy metal ions by various low-cost adsorbents: a review. Int. J. Environ. Anal. Chem. 102(2):342–379. https://doi.org/10.1080/03067319.2020.1722811

[4] Tan X., Wei H., Zhou Y., Zhang C., Ho S.-H. (2022) Adsorption of sulfamethoxazole via biochar: the key role of characteristic components derived from different growth stage of microalgae. Environ. Res. 210:11296. https://doi.org/10.1016/j.envres.2022.112965

[5] Elmoutez S.S., Abushaban A., Necibi M.C., Sillanp M., Liu J., Dhiba D., Chehbouni A., Taky M. (2023) Design and operational aspects of anaerobic membrane bioreactor for efficient

wastewater treatment and biogas production. Environ. Challenges 10:100671. https://doi.org/10.1016/j.envc.2022.100671

[6] Lanzetta A., Papirio S., Oliva A., Cesaro A., Pucci L., Capasso E.M., Esposito G., Pirozzi F. (2023) Ozonation processes for color removal from urban and leather tanning wastewater. Water 15(13):2362. https://doi.org/10.3390/w15132362

[7] Ang W.L., Mohammad A.W. (2020) State of the art and sustainability of natural coagulants in water and wastewater treatment. J. Clean. Prod. 262:121267. https://doi.org/10.1016/j.jclepro.2020.121267

[8] Renault F., Sancey B., Badot P.M., Crini G. (2009) Chitosan for coagulation/flocculation processes an eco-friendly approach. Eur. Polym. J. 45(5):1337–1348. https://doi.org/10.1016/j.eurpolymj.2008.12.027

[9] El-Gaayda J., Titchou F.E., Oukhrib R., Yap P.-S., Liu T., Hamdani M., Ait Akbour R. (2021) Natural flocculants for the treatment of wastewaters containing dyes or heavy metals: A state-of-the-art review. J. Environ. Chem. Eng. 9(5):106060. https://doi.org/10.1016/j.jece.2021.106060

[10] Metin S., ifi D.. (2023) Chemical industry wastewater treatment by coagulation combined with Fenton and photo-Fenton processes. J. Chem. Technol. Biotechnol. 98(5):1158–1165. https://doi.org/10.1002/jctb.7321

[11] Ahmad A, Abdullah S.R.S., Hasan H.A., Othman A.R., Ismail N.I. (2021) Plant-based versus metal-based coagulants in aquaculture wastewater treatment: Effect of mass ratio and settling time. J. Water Process Eng. 43:102269. https://doi.org/10.1016/j.jwpe.2021.102269

[12] Kurniawan S.B., Abdullah S.R.S., Imron M.F., Said N.S.M., Ismail N., Hasan H.A., Othman A.R., Purwanti I.F. (2020) Challenges and opportunities of biocoagulant/bioflocculant application for drinking water and wastewater treatment and its potential for sludge recovery. Int. J. Environ. Res. Public Health 17(24):9312. https://doi.org/10.3390/ijerph17249312

[13] Gautam S., Saini G. (2020) Use of natural coagulants for industrial wastewater treatment.Glob. J. Environ. Sci. Manag. 6(4):553–578. https://doi.org/10.22034/gjesm.2020.04.10

[14] Ayangbenro A.S., Babalola O.O., Aremu O.S. (2019) Bioflocculant production and heavy metal sorption by metal resistant bacterial isolates from gold mining soil. Chemosphere 231:113–120. https://doi.org/10.1016/j.chemosphere.2019.05.092

[15] Pu L., Zeng Y.J., Xu P., Li F.Z., Zong M.H., Yang J.G., Lou W.Y. (2020) Using a novel polysaccharide BM2 produced by Bacillus megaterium strain PL8 as an efficient bioflocculant for wastewater treatment. Int. J. Biol. Macromol. 162:374–384. https://doi.org/10.1016/j.ijbiomac.2020.06.167

[16] Giri S.S., Harshiny M., Sen S.S. (2015) Production and characterization of a thermostable bioflocculant from Bacillus subtilis F9, isolated from wastewater sludge. Ecotoxicol. Environ. Saf. 121:45–50. https://doi.org/10.1016/j.ecoenv.2015.06.010

[17] Sivasankar P., Poongodi S., Lobo A.O., Pugazhendhi A. (2020) Characterization of a novel polymeric bioflocculant from marine actinobacterium Streptomyces sp. And its application in recovery of microalgae. Int. Biodeterior. Biodegrad. 148:104883. https://doi.org/10.1016/j.ibiod.2020.104883

[18] Yafetto L. (2022) Application of solid-state fermentation by microbial biotechnology for bioprocessing of agro-industrial wastes from 1970 to 2020: A review and bibliometric analysis. Heliyon 8(3). https://doi.org/10.1016/j.heliyon.2022.e09173

[19] Selo G., Planini M., Tima M., Tomas S., Koceva Komleni D., Buci-Koji A. (2021) A comprehensive review on valorization of agro-food industrial residues by solid-state fermentation. Foods 10(5):927. https://doi.org/10.3390/foods10050927

[20] Farinas C.S. (2015) Developments in solid-state fermentation for the production of biomass-degrading enzymes for the bioenergy sector. Renewable Sustainable Energy Rev. 52:179–188. https://doi.org/10.1016/j.rser.2015.07.092

[21] Prado-Barragn L.A., Figueroa J.J.B., Rodrguez Durn L.V., Aguilar Gonzlez C.N., Hennigs C. (2016) Fermentative production methods. In: Biotransformation of Agricultural Waste and By-Products, pp. 189–217. Elsevier, Amsterdam. https://doi.org/10.1016/B978-0-12-803622-8.00007-0

[22] Guo J., Yang C., Peng L. (2014) Optimization of microbial flocculant production by response surface methodology. Bioresour. Technol. 152:490–498. http://10.1007/s12088-016-0631-3

[23] Fellah M., Alam M.Z., Al-Mamun A., Khabbashi N.A., Engliman N.S., Arab S.H. (2023) Effect of the lignocellulolytic substrates and fermentation process parameters on microbial coagulant production for water treatment. IIUM Eng. J. 24(1):13–26. https://doi.org/10.31436/iiumej.v24i1.2400

[24] Fellah M., Alam M., Al-Mamun A., Kabbashi N., Engliman N. (2021) Evaluation of solid-state production of microbial coagulant using various lignocellulosic media to reduce water turbidity. IOP Conf. Ser.: Mater. Sci. Eng. 1192(1):012022. https://doi.org/10.1088/1757-899X/1192/1/012022

[25] Maran J.P., Manikandan S., Nivetha C.V., Dinesh R. (2017) Ultrasound assisted extraction of bioactive compounds from Nephelium lappaceum L. fruit peel using central composite face centered response surface design. Arabian J. Chem. 10:1145–1157. https://doi.org/10.1016/j.arabjc.2013.02.007.\

[26] Zulkeflee Z., Snchez A. (2014) Solid-state fermentation of soybean residues for bioflocculant production in a pilot-scale bioreactor system. Water Sci. Technol.70(6):1032–1039. https://doi.org/10.2166/wst.2014.329

[27] Aljuboori, A. H. R., Uemura, Y., Osman, N. B., & Yusup, S. (2014). Production of a bioflocculant from Aspergillus niger using palm oil mill effluent as carbon source. Bioresource technology, 171,66-70. http://dx.doi.org/10.1016/j.biortech.2014.08.038.

[28] Ma, C. H., Liu, T. T., Yang, L., Zu, Y. G., Chen, X., Zhang, L., ... & Zhao, C. (2011). Ionic liquid-based microwave-assisted extraction of essential oil and biphenyl cyclooctene lignans from Schisandra chinensis Baill fruits. Journal of Chromatography a, 1218(48), 8573-8580. https://doi.org/10.3390/ijms131114294

[29] David, O.; Oluwole, O.; Ayodele, O.; Lasisi, T. Characterisation of fungal bioflocculants and its application in water treatment. Curr. J. Appl. Sci. Technol. 2019, 34, 1 9. http://10.9734/cjast/2019/v34i630159

[30] Tsilo, P. H., Basson, A. K., Ntombela, Z. G., Maliehe, T. S., & Pullabhotla, R. V. (2021). Isolation and optimization of culture conditions of a bioflocculant-producing fungi from Kombucha tea SCOBY. Microbiology Research, 12(4), 950-966. https://doi.org/10.3390/microbiolres12040070

[31] HASSAN, H. B., ALAM, M. Z., Al Mamun, A. B. D. U. L. L. A. H., & MUJELI, M. (2021). Screening of Nutrients for The Production of Microbial coagulant from Lentinus Squarrosulus for Water Treatment. Evaluation, 1, 12.

[32] Pu, S., Ma, H., Deng, D., Xue, S., Zhu, R., Zhou, Y., & Xiong, X. (2018). Isolation, identification, and characterization of an Aspergillus niger bioflocculant-producing strain using potato starch wastewater as nutrilite and its application. PloS one, 13(1), e0190236. https://doi.org/10.1371/journal.pone.0190236

[33] Ramachandran, S., Patel, A. K., Nampoothiri, K. M., Francis, F., Nagy, V., Szakacs, G., & Pandey, A. (2004). Coconut oil cake—a potential raw material for the production of α-amylase. Bioresource technology, 93(2), 169-174. https://10.1016/j.biortech.2003.10.021

[34] Krishna, P. R., Srivastava, A. K., Ramaswamy, N. K., Suprasanna, P., & D'souza, S. F. (2012). Banana peel as substrate for alpha-amylase production using Aspergillus niger NCIM 616 and process optimization. Indian journal of biotechnology, 11(3), 314-319.

https://doi.org/10.31436/iiumej.v27i2.3843

Eco-Friendly Asphalt Mixture: Utilization of Waste Polypropylene as a Partial Replacement of Aggregate

ISRAA SHAMKY DARYUL, ROAA HAMED LATIEF*

Department of Civil Engineering, College of Engineering, University of Baghdad, Baghdad, Iraq

**Corresponding author: roaa.hamed@coeng.uobaghdad.edu.iq*

(Received: 11 July 2025; Accepted: 23 January 2026; Published online: 10 May 2026)

ABSTRACT: Recently, many techniques have been developed to improve the properties of asphalt pavement and to reduce environmental pressure. Adding waste polypropylene (WPP) to asphalt mixtures is one of the most often used and effective techniques. Nevertheless, waste plastic aggregate in flake form exhibited the worst performance in the asphalt mixture, as demonstrated in previous investigations. Therefore, a new recycling process is used in this research to produce crushed and angular WPP aggregate (WPPA) with controlled dimensions, usually ranging from 0.075 mm to 12.5 mm, similar to those of natural aggregate. WPPA is used as a partial replacement for coarse aggregates, fine aggregates, and mineral filler, replacing all aggregate sizes from 12.5 mm to less than 0.075 mm. Modified asphalt mixture was designed with replacement proportions of 4%, 6%, and 8% for coarse WPPA; 2%, 4%, and 6% for fine WPPA; and 0.5%, 0.75%, and 1.0% for filler WPPA by the weight of the asphalt mixtures. An abrasion test was performed to assess the resistance and durability of WPPA to tearing and wear. Marshall and indirect tensile strength tests were employed to assess the performance of control and WPP-modified mixtures. Compared to the control mixture, WPPA incorporation provided greater stability and higher stiffness. For instance, notable increases in stability (up to 34.8%) and stiffness (up to 42.8%) were observed upon replacing fine aggregates with 2% WPPA in the asphalt mixture. In addition, a significant enhancement in the resistance of WPPA-modified mixture against moisture damage was achieved as the tensile strength ratio (TSR) was increased substantially, especially at replacement ratios of 4% coarse WPPA, 2% fine WPPA, and 0.75% filler WPPA, where TSR was increased by 14.4%, 10.9%, and 8.8%, respectively. In conclusion, WPPA-modified mixtures are less susceptible to moisture damage with higher stability and stiffness.

ABSTRAK: Dalam beberapa tahun terakhir, banyak teknologi telah dikembangkan untuk meningkatkan sifat-sifat pengaspalan asfalt dan mengurangkan beban alam sekitar . Penambahan sisa polipropilena (WPP) kepada campuran asfalt adalah salah satu teknik yang paling kerap digunakan dan berkesan. Dalam kajian ini, agregat WPP (WPPA) digunakan sebagai penggantian separa agregat kasar, agregat halus, dan pengisi mineral, menggantikan semua kecerunan agregat daripada saiz 12.5 mm kepada saiz kurang daripada 0.075 mm. Kompaun plastik terpakai dalam bentuk kepingan menunjukkan prestasi yang lebih buruk untuk campuran aspal seperti yang dibuktikan dalam penyelidikan sebelum ini. Oleh itu, satu proses kitar semula baru sedang digunakan untuk menghasilkan WPPA dengan dimensi yang terkawal, biasanya antara 0.075 mm hingga 12.5 mm. Campuran asfalt diubah suai telah direka bentuk dengan perkadaran gantian 4%, 6%, dan 8% untuk WPPA kasar; 2%, 4% dan 6% untuk WPPA halus; dan 0.5%, 0.75% dan 1.0% untuk pengisi WPPA mengikut berat campuran asfalt berat. Ujian lelasan dilakukan untuk menilai rintangan dan ketahanan WPPA untuk koyak dan haus. Marshall dan ujian kekuatan tegangan tidak langsung telah digunakan untuk menilai prestasi campuran kawalan dan campuran asfalt yang diubah suai dengan WPP. Berbanding dengan campuran kawalan, penggabungan WPPA memberikan kestabilan yang lebih tinggi dan kekukuhan yang lebih baik, yang masing-masing meningkat sebanyak 7-35%

https://doi.org/10.31436/iiumej.v27i2.3843

dan 1-43%. Menurut ujian kekuatan tidak langsung, peningkatan ketara dalam rintangan campuran diubah suai WPPA terhadap kerosakan lembapan telah dicapai kerana nisbah kekuatan tegangan (TSR) telah meningkat dengan ketara terutamanya pada nisbah penggantian 4% WPPA kasar, 2% WPPA halus dan 0.75% pengisi WPPA. Secara kesimpulannya, campuran yang diubah suai WPPA kurang terdedah kepada kerosakan lembapan dengan kestabilan dan kekakuan yang lebih tinggi.

KEYWORDS: *Waste polypropylene, Recycling process, Aggregate replacement, Dry process, Asphalt binder.*

1. INTRODUCTION

In recent decades, large quantities of plastic waste have been generated worldwide, negatively affecting the environment. In some regions, such as European countries, approximately 30% of total plastic waste is recycled, while the remaining 70% is disposed of in landfills [1]. Globally, around 9% of waste plastic was recycled, while the remaining quantities were landfilled, leaked into the environment, were mismanaged, or were incinerated [2]. As the push for sustainability grows stronger, scientists are increasingly exploring innovative approaches to recycling and repurposing waste plastic. Therefore, pavement researchers have extensively investigated the use of various types of plastic in asphalt mixtures as a binder modifier, additive, partial replacement for aggregates, or a combination of these [3].

Currently, polymer-modified asphalt mixtures achieve more desirable pavement performance, enabling them to resist heavy loads and harsh environmental conditions. For instance, modifying the asphalt binder with 6% high-density polyethylene increased stability by approximately 113% and improved resistance to water damage by more than 10% in modified asphalt mixtures [4]. In addition, incorporating waste plastic in asphalt pavement contributes to sustainability by lowering costs and reducing pollution. For clarification, using waste plastic after recycling in the pavement sector can help save energy, preserve natural resources, lower harmful emissions, and reduce landfill waste. The first use of waste plastics in the pavement field originated in the 1990s, primarily in fiber form to modify the properties of asphalt mixtures [5-6].

In fact, polymers can be categorized into two groups: elastomers, such as styrene-butadiene rubber, polybutadiene, and styrene-butadiene-styrene, which enhance both elasticity and strength, and plasterers, such as polyethylene, polypropylene, and ethylene vinyl acetate, which only enhance strength [7]. Polypropylene is the second-most-produced plastic globally, accounting for 21% of the total plastic market [8]. Its widespread use highlights its importance in various industries. Polypropylene, discovered in 1951, gained rapid popularity due to its low density among commodity plastics. It offers excellent chemical resistance and can be processed using various methods, such as injection molding and extrusion, making it highly versatile and widely used in packaging, medical supplies, automotive components, housewares, and consumer goods [9].

Hot-mix asphalt primarily consists of aggregates and asphalt binder. Thus, incorporating waste plastic into asphalt mixtures as a partial replacement for aggregates reduces dependence on traditional virgin aggregates, helping conserve vital natural resources. Waste plastic is added using a dry process, meaning these wastes are incorporated into the aggregate prior to encountering the binder. This method forms a thin plastic coating around the aggregate, and when the coated aggregate is combined with the binder, it delivers outstanding performance in the asphalt mixture because polymers typically have higher viscosity than asphalt binders and

better adhesion to the aggregate [10]. In addition, chemical interactions can form between high-viscosity polymers and polar components on the natural aggregate surface, thereby forming an additional adhesion layer. Thus, the mechanical interlocking of the aggregates, internal strength, cohesion, and the resistance of asphalt pavement to water stripping were improved [11, 12]. In general, the degree of interaction of the polymers with any material depends on their softening temperature, as well as on the time of digestion and mixing [10].

In this research, waste polypropylene (WPP) is used as a partial replacement of aggregates in an asphalt mixture. Incorporating WPP-modified aggregate (WPPA) into asphalt mixtures is an effective and practical option, not only from an environmental perspective but also from an economic perspective, since aggregates comprise nearly 95% of the asphalt mixture. Generally, incorporating WPPA as an artificial aggregate in asphalt mixtures is an economical choice due to reduced energy consumption, lower waste disposal costs, decreased transportation needs, lower material costs, and improved pavement performance, leading to long-term savings [13]. Several studies have examined the use of waste plastic as a partial replacement for natural aggregates in asphalt mixtures; however, WPPA has been used in only a limited number of investigations. Thus, it is imperative to highlight the significance and originality of this study on environmentally friendly asphalt mixtures made from waste plastic.

In fact, the dry process is more appropriate for incorporating WPP into the asphalt mixture, since WPP has a melting point above 160 °C. In turn, WPP faces difficulties during blending with the binder via the wet process because its melting point is higher than the binder's mixing temperature; thus, it is difficult to melt completely within the binder structure, leading to premature binder aging and making the pre-mixing process difficult [14]. The asphalt binder becomes brittle and stiff due to premature aging during the modification process with plastic, leading to a severe reduction in pavement service life by increasing its susceptibility to various distresses, particularly cracking [15]. Whereas WPP was added into asphalt mixtures in a dry process as an artificial aggregate at typical mixing temperatures, avoiding separation and melting issues. In addition, the dry process is more cost-effective than the wet process because it eliminates the need for specialized equipment and pre-melting [16].

Hence, Kim et al. [17] studied the incorporation of waste plastic aggregate, including WPPA, as a partial replacement of natural coarse aggregate at 10% and 20% by aggregate weight. The WPPA sizes ranged between 5 and 13 mm. The outcomes revealed that the WPPA-mixture exhibited a 23% higher Marshall stability than the control mixture. Also, the tensile strength of the WPPA mixture increased by 20% and 18% for dry and wet conditions, respectively. Whereas Bumruiana et al. [18] used WPP, at sizes ranging from 0.1 mm to 2 mm, as a mixture modifier, adding it at three proportions: 0.1%, 0.3%, and 0.6% by weight of the mixture. Adding 0.3% WPP raised the Marshall stability of the asphalt mixture by 32% compared to the control mixture.

Lee et al. [19] opted to use WPP in a fine size (less than 5 mm) as an artificial aggregate in lower pavement layers, which was added at 3%, 5%, 7%, and 10% to conserve natural resources of natural aggregate and to reduce pressure on the environmental, it was combined with the addition of magnesium powder (particle size less than 0.075 mm) to improve the performance of the mixture. The results reveal that the asphalt mixture with 5% fine WPPA demonstrated excellent water damage resistance, greater strength, and higher resistance to plastic deformation, with improvement rates of 31%, 39%, and 60%, respectively, compared to the control mixture prepared with natural aggregates.

Dione et al. [20] used 50% WPPA and 50% waste low-density polyethylene as partial replacements for natural aggregate, with contents ranging from 2% to 8%. Also, these waste

plastics are incorporated into asphalt mixtures in a flake shape with a fine size between 2 mm and 5 mm. The research showed that asphalt mixture performance improved with the addition of waste plastic at percentages below 7%, with increased Marshall stability, enhanced moisture damage resistance, and an 80% reduction in rut depth compared to the mixture without waste plastic. This performance enhancement can be achieved through two main mechanisms. The first mechanism is to improve asphalt-aggregate adhesion by effectively coating the aggregate surfaces. The second mechanism uses WPPA as a reinforcement material within the aggregate skeleton [21]. However, in other studies, using waste plastic in flake form as a partial replacement for aggregate had a negative impact on the mixture's performance, contributing to lower permanent deformation, indirect tensile strength, stiffness, and moisture damage resistance [22-23].

Although promising outcomes have been obtained, these studies have often overlooked crucial gaps and shortcomings. The significant gap is the limited use of WPPA as a partial replacement for aggregates in asphalt mixtures at different ratios. In addition, the literature lacks sufficient in-depth research examining the effects of varying WPPA sizes and angular shapes on asphalt mix performance. Researchers have observed that the irregular structure, sharp edges, and higher surface area of angular aggregates can improve mechanical interlocking and bonding between aggregate particles within the asphalt mixture more than those of rounded aggregates and flake-shaped plastic aggregates. This mechanical interlocking is crucial for forming a strong aggregate skeleton and effective load transfer, leading to higher stability, better workability, higher shear strength, and superior resistance to deformation, particularly under elevated temperatures and heavy traffic loads [24, 25].

These gaps motivate this research, which aims to systematically assess and enhance WPPA to increase the sustainability and viability of asphalt mixtures. The aim of this research was to assess the feasibility of replacing natural aggregates with WPPA using a dry process, without compromising the performance of the asphalt mixture while preserving natural aggregate resources. A new recycling process is employed to produce WPPA with different sizes, controlled dimensions, and angular shapes. Three WPPA sizes (coarse, fine, and filler) were considered. Pavement performance was evaluated using a variety of tests, including abrasion, Marshall, and water damage resistance tests.

2. MATERIALS

2.1. Asphalt Binder

Asphalt binder Type (40–50) is sourced from the Al-Dora refinery oil in Baghdad, Iraq, and is commonly used in local paving works because it has relatively high hardness and stiffness compared to other types, such as 50-60 and 60-70. Thus, it is better suited for withstanding high-traffic loads and hot climates, especially in the middle and southern regions of Iraq. Laboratory tests were conducted in accordance with ASTM standards, and the results were compared with Iraqi road and bridge specifications [26]. In this study, asphalt binder (Type 40-50) was utilized to produce both conventional and modified asphalt mixtures. Before its application in mixtures, the physical properties of the asphalt binder, including specific gravity, ductility, softening point, flash point, and penetration, were thoroughly analyzed, as shown in Table 1. Also, the penetration and ductility tests were repeated on the residue from the thin-film oven test to assess changes in the asphalt binder's flexibility and consistency at high temperatures during production and construction. Correspondingly, this can help predict the pavement's performance, particularly in terms of durability and cracking resistance.

https://doi.org/10.31436/iiumej.v27i2.3843

Table 1. Physical properties of asphalt binder

Property	Test Method	Results	Requirements [26]
Penetration, 1/10 mm	ASTM D-5	43	40-50
Specific Gravity, unit-less	ASTM D-70	1.04	-
Ductility, cm	ASTM D-113	133	≥100
Flash Point, ºC	ASTM D-92	324	232 min
Softening Point, ºC	ASTM D-36	51	-
Thin Film Oven Test (TFOT)	ASTM D-1754	-	-
Retained Penetration after TFOT, %	ASTM D-5	75	> 55
Ductility after TFOT, cm	ASTM D-113	95	> 25

2.2. Aggregates

The fine and coarse aggregates utilized in this study were sourced from the Al-Nibaai quarry. Coarse aggregates were defined as those retained between sieves No. 3/4 and No. 4, while fine aggregates were classified as those retained between sieves No. 4 and No. 200. In the road laboratory, these aggregates were sieved and blended to conform to the local specification [26]. In addition, limestone dust, obtained from a local factory, was used as the mineral filler in this study. Generally, this filler is used in conventional asphalt mixtures. The physical properties of the filler were examined, revealing a bulk specific gravity of 2.70 and a passing percentage of 88% through sieve No. 200. Physical tests were performed to evaluate the properties of both fine and coarse aggregates, and their results are presented in Table 2. In addition, Fig. 1 shows the selected gradation of aggregates, including coarse, fine, and filler materials, for conventional asphalt mixtures used in the surface layer, which aligns with the criteria limits. These limits were stipulated by the local specification [26].

Table 2. Physical properties of coarse and fine aggregates

Property	Coarse Aggregate		Fine Aggregate	
	Results	Test Method	Results	Test Method
Bulk Specific Gravity	2.58	ASTM C-127	2.6	ASTM C-128
Apparent Specific Gravity	2.61	ASTM C-127	2.72	ASTM C-128
Water Absorption, %	1.024	ASTM C-127	1.731	ASTM C-128
Clay Lumps, %	0.509	ASTM C-142	0.98	ASTM C-142
Passing Sieve No. 200, %	0.46	ASTM C-117	0.94	ASTM C-117

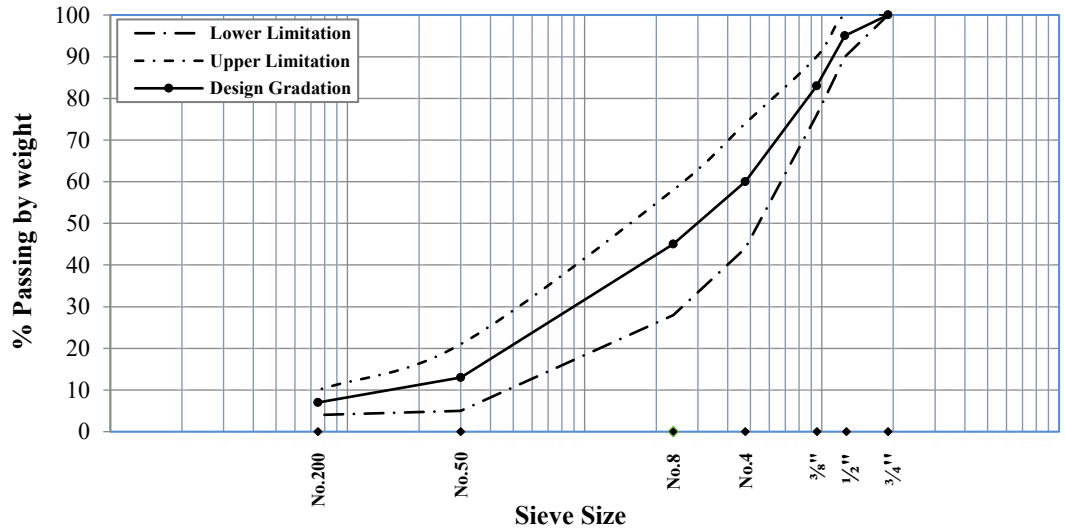


Figure 1. Aggregate gradation for the surface layer.

https://doi.org/10.31436/iiumej.v27i2.3843

2.3. Waste Polypropylene

In this work, WPP was sourced from local disposable plates, food packaging, microwave food containers, automotive parts, and cups. ASTM C128 was followed and used to measure the water absorption of WPP. The physical characteristics of WPP are listed in Table 3.

Table 3. Physical properties of WPP

Property	Value
Color	Transparent
Melting Point, °C	150-165
Water Absorption (%)	1.3

3. EXPERIMENTAL METHODS

3.1. Production of WPP Aggregates

Figure 2 illustrates the production stages of WPPA. Firstly, WPP undergoes manual sorting (Fig. 2a). The sorted plastics are then cleaned thoroughly, shredded into small flakes, and soaked and dried to eliminate dirt and dust, as shown in Fig. 2b. The plastic resin is melted at temperatures of 160–180 °C (Fig. 2c) [17, 19], transforming it into a molten form ready for further processing, as presented in Fig. 2d. Following that, molten WPP is carefully extruded through a custom-designed die, producing WPPA with controlled dimensions, typically ranging from 12.5 mm to 0.075 mm (Fig. 2e). These WPPA have sharp edges and an irregular surface similar to the shape of natural aggregate particles. To use WPP as a filler material, the molten WPP is cooled to room temperature, collected in a flat pan, and mechanically ground for approximately 5 min into a fine powder with particle sizes below 0.075 mm, as illustrated in Fig. 2f.

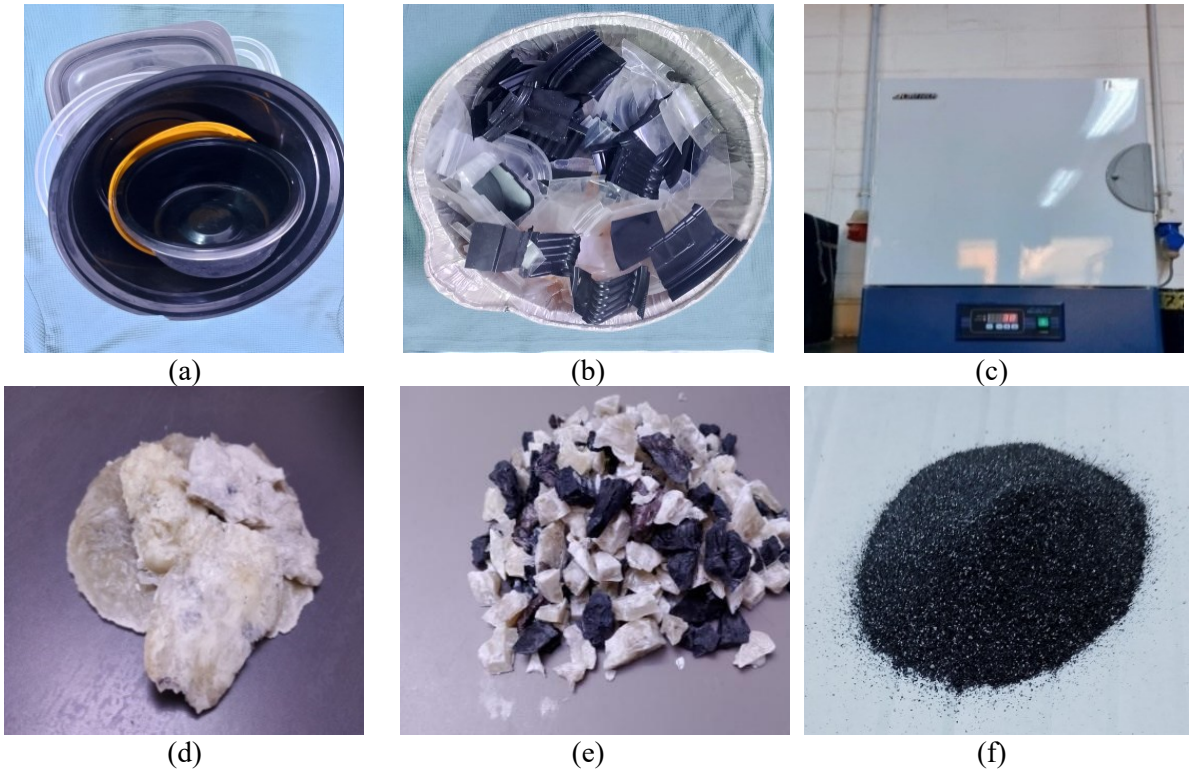


Figure 2. Manufacturing WPPA: (a) collection of WPP; (b) shredded WPP; (c) melted WPP; (d) cooled WPP; (e) WPPA in coarse and fine sizes; and (f) WPPA powder.

https://doi.org/10.31436/iiumej.v27i2.3843

This process ensures consistency in size and quality, making WPPA suitable for use in construction roads. In this study, the replacement proportions (by the weight of asphalt mixture) were used as follows: 4%, 6%, and 8% for coarse WPPA; 2%, 4%, and 6% for fine WPPA; and 0.5%, 0.75%, and 1% for filler WPPA. The replacement contents of WPPA for fine and coarse aggregates were selected based on findings from previous studies [19, 20, 27]. Whereas the replacement contents of WPPA for the filler size have been selected by the authors, no previous studies have replaced the filler materials with recycled plastic waste. The natural aggregates and WPPA sizes are depicted in Fig. 3.

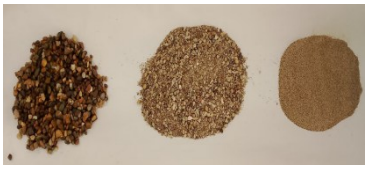
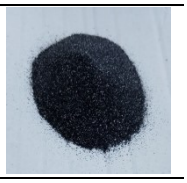
Type of Aggregate	**Type of Particle**						
	Coarse			Fine			Filler
Natural Aggregate							
WPPA							
Particle Size, mm	12.5	9.5	4.75	2.36	0.3	0.075	Less than 0.075

Figure 3. Surface appearance and sizes of natural aggregate and WPPA.

3.2. Abrasion Test

One of the goals of this research was to partially substitute natural aggregates retained on sieve sizes (12.5-9.5 mm) with WPPA at replacement proportions of 4, 6, and 8%. The Los Angeles test was used to calculate aggregate wear resistance, and the abrasion loss value should be less than 30% for pavements, as specified by the local specification for roads and bridges [26]. The Los Angeles abrasion test, following the ASTM C-131 standard, is presented in Fig. 4. This test is performed to evaluate the resistance of a mixture of WPPA and natural aggregate to abrasion and impact. There are six types of aggregate grading (A, B, C, D, E, and F) in the ASTM C-131 standard. Among these types, grading type B is more appropriate for local aggregate gradation for the surface layer, as specified by the local specification [26]. For grading type B, a sample of coarse aggregate at a weight of 5000 gm (2500 gm retained on a sieve with a size opening of 12.5 mm (½ inch) and 2500 gm retained on a sieve with a size opening of 9.5 mm (⅜ inch)) is combined with eleven steel spheres and placed inside a rotating steel drum. The drum rotates for a designated number of revolutions (typically 500 for grading type B), subjecting the aggregate to continuous crushing, abrasion, and grinding. After completing the rotation, the aggregate is removed, sieved through a No. 12 sieve, and the weight retained on the sieve is measured. The percentage of material loss, known as the Los Angeles abrasion or abrasion loss, is calculated using the formula:

$$Abrasion\ Loss\ (\%) = \frac{W_{Original} - W_{Final}}{W_{Original}}\ x\ 100 \quad (1)$$

where $W_{Original}$ is the original weight of the aggregate sample before the Los Angeles test in gm, and W_{Final} is the retained weight of the aggregate sample after the Los Angeles test in grams.

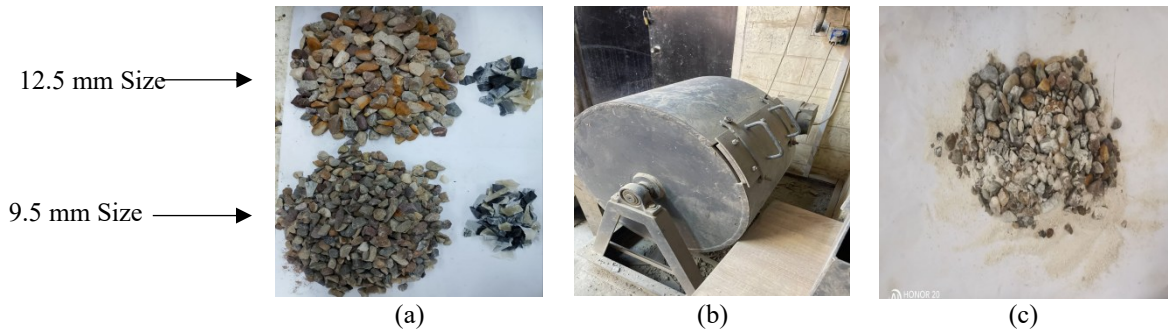


Figure 4. Los Angeles abrasion test: (a) sample of natural aggregate and WPPA; (b) Los Angeles drum; and (c) sample after test.

3.3. Asphalt Mixture Sample Preparation

The established combinations total 102 samples. Marshall samples, measuring 101.6 mm in diameter and 63.5 mm in height, are used in asphalt mixture design tests to assess key properties. Firstly, density and air voids (AV) were measured to evaluate the mixture compaction and to determine the percentage of air-filled spaces, respectively, which mainly influence durability. Secondly, voids in mineral aggregate (VMA) were computed to assess the asphalt film thickness on aggregate surfaces, which is vital for durability. Whereas voids filled with asphalt (VFA) were calculated to determine the percentage of air voids filled with asphalt, which is crucial for achieving the desired balance among permeability, compaction, and durability in the asphalt mixture. Thirdly, Marshall stability was measured to evaluate the strength of the asphalt mixture, which indicates load capacity. At the same time as measuring Marshall stability, the flow was recorded to assess deformation under load, which represents deformation prior to failure. Finally, the resistance of asphalt mixtures (modified and unmodified) to moisture (water) was assessed.

Table 4. Number of samples for control and WPPA-modified mixtures distributed based on Marshall and water damage tests

Mixture Type		Replacement Proportions of WPPA (%), by the Weight of Asphalt Mixture	Sample Number		
			Marshall Test	Water Damage Test	
				Dry Condition	Wet Condition
Control Mixture		-	15	3	3
Modified Mixture	WPPA Coarse	4	3	3	3
		6	3	3	3
		8	3	3	3
	WPPA Fine	2	3	3	3
		4	3	3	3
		6	3	3	3
	WPPA Filler	0.50	3	3	3
		0.75	3	3	3
		1.00	3	3	3
Total Number of Samples			102 sample		

For the control mixture, the aggregate was blended with asphalt and stirred for three minutes. This ensured that the aggregate was thoroughly coated by the binder, achieving uniform coverage [28]. For mixtures modified with WPP, the mix of natural aggregates and WPPA was placed in an oven for three to fifteen minutes at a temperature of 170-180 °C [29]. It formed a uniform coating around the aggregates after hot bitumen at 160°C was incorporated

https://doi.org/10.31436/iiumej.v27i2.3843

into the mixture. The samples were compacted to 75 blows per side for the Marshall test. Meanwhile, for water damage test samples, 54 blows per side are applied to achieve the required air voids. After compaction, samples were left to cure in the molds for 24 hours at room temperature. Table 4 provides the total number of samples by mixture type, the WPPA replacement ratio for the modified mixture, and the number of tests conducted.

3.4. Marshall Test

According to ASTM D6927, the Marshall mix design is commonly used in Iraq to optimize hot asphalt mixtures. The asphalt content for hot mixtures used in surface and binder layers was specified as 4%-6% in the local criteria [26]. Therefore, these mixtures were designed using five asphalt content levels: 4.0%, 4.5%, 5.0%, 5.5%, and 6.0%. A total of 15 specimens were prepared, three for each asphalt percentage, to determine the optimal asphalt content (OAC). The Asphalt Institute's method was employed to calculate the OAC that achieved maximum density, maximum stability, and 4% air voids. For each sample, the bulk density, AV, VMA, and VFA were determined. Following this, the samples were conditioned at 60 °C for 30-40 minutes. These samples were tested using the Marshall loading device at a rate of 50.8 mm/min until the maximum load was reached, which indicates Marshall stability. At that load, the flow is measured, which indicates the deformation in mm. The relationship between flow and stability represents a delicate balance and trade-off between the asphalt mixture's flexibility and its ability to resist deformation. For the selected gradation, the OAC was found to be 5% for the control mixture, and it was also used for WPPA-modified mixtures.

3.5. Indirect Tensile Test

The indirect tensile test is a widely used method for evaluating the water-damage performance of flexible pavements. In this research, sixty specimens were prepared and tested under two conditions: dry and wet. Six specimens represented a control mixture with no WPPA addition, while fifty-four specimens were prepared for WPPA-modified mixtures.

The stages of performing this test are presented in Fig. 5. Specimens were compacted using an automatic Marshall impact compactor to achieve a target air void of 7±1%. The test was conducted in accordance with ASTM D 4867 M. Specimens were divided into two groups: unconditioned and conditioned. For conditioned specimens, water saturation was achieved using vacuum desiccators, with a saturation level between 55% and 80%. Saturated samples were then wrapped in stretch film, placed in plastic bags containing 10 ml of water, and stored at -18°C for 16 hours. Afterward, these samples were moved to a water bath at 60 °C for 24 hours. Unconditioned samples were simply submerged in a water bath for 20 minutes, while conditioned samples were immersed in a 25°C water bath for one hour prior to testing. The samples were subjected to an indirect tensile splitting test in a vertical orientation to assess their resistance to water damage. Tensile Strength Ratio (TSR) of asphalt mixtures was determined using the formula:

$$TSR\ (\%) = \frac{ITS_{condition}}{ITS_{uncondition}}\ x\ 100 \tag{2}$$

$$ITS = \frac{2000 * P}{\pi\ t\ D} \tag{3}$$

where ITS is the indirect tensile strength in kPa; P indicates the maximum load to fracture the sample in N; t refers to the thickness of the sample in mm, and D is the sample diameter in mm.

https://doi.org/10.31436/iiumej.v27i2.3843



Figure 5. Stages of water damage test: (a) samples before test; (b) saturation process; (c) freezing cycle; (d) thawing cycle; (e) sample testing; and (f) fractured samples.

4. Results and Discussion

4.1. Abrasion Resistance

Figure 6 presents the results of abrasion loss for both mixtures. The results reveal that the abrasion loss was reduced by approximately 10.2% in the sample with 8% WPPA compared with the natural aggregate sample. As coarse WPPA content increased, the aggregate proportion decreased equivalently, leading to reduced wear and corrosion. The higher polymer content results in a lower percentage of abrasion loss. The WPP is classified as a thermoplastic polymer, and it has a plastic nature. Thus, after the recycling process, WPP retains high strength, preventing its particles from shattering or breaking during the abrasion test. In addition, replacing portions of natural aggregate with WPPA reduces crushing between the steel balls and the natural aggregate particles because WPPA partially surrounds the aggregate particles. This improvement in aggregate abrasion resistance is similar to that reported in a

previous study [30], which used polyethylene as a partial replacement for aggregates. Also, all samples fulfilled the maximum requirement (30%).

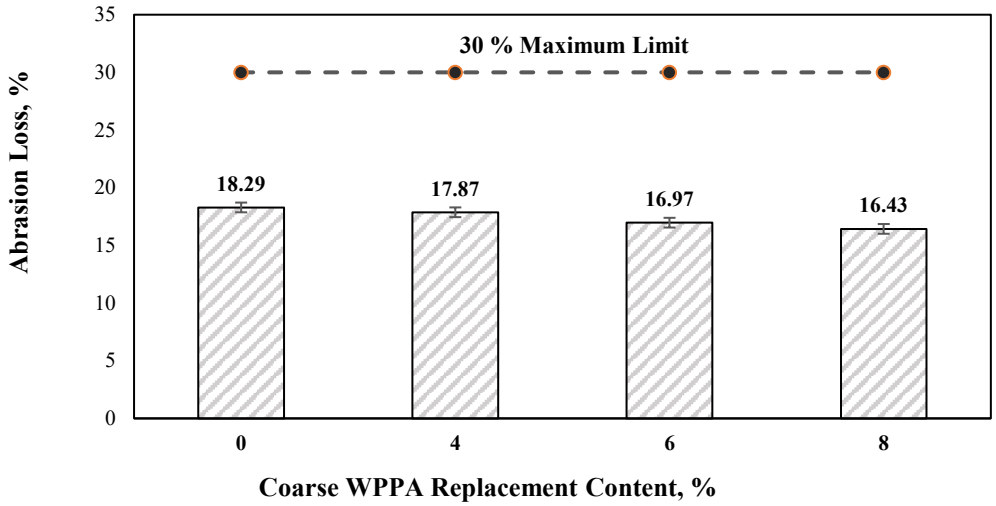


Figure 6. Abrasion loss results of natural aggregate and WPPA.

4.2. Marshall Characteristics

4.2.1. Mechanical Properties

As observed in Fig. 7, most modified mixtures with WPPA showed higher stability values than the control mixture. In addition, WPPA, in coarse, fine, and mineral filler sizes, enhances the stability of the asphalt mixture due to improved aggregate interlock, but excessive amounts reduce the stability value. More precisely, the stability at the addition of 6% fine WPPA and 8% coarse WPPA was reduced by approximately 21% and 33%, respectively, compared to the addition of 2% fine WPPA and 4% coarse WPPA.

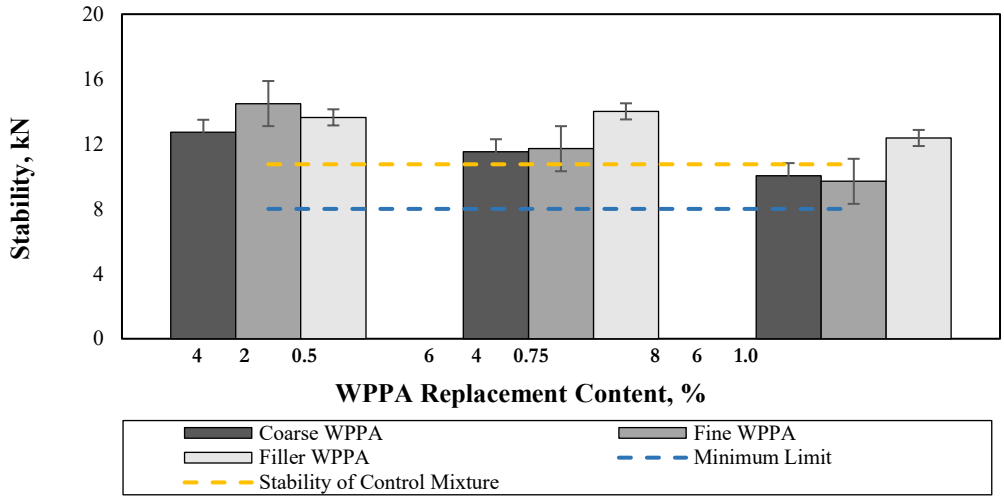


Figure 7. Marshall stability of control and WPPA-modified mixtures.

The drop in stability occurs because WPPA has a lower specific gravity, which weakens aggregate bonding, especially at higher concentrations. To clarify, the overall bulk density of the asphalt mixture decreases when portions of natural aggregate are replaced with WPPA, resulting in a higher air void percentage in the compacted mixture. In addition, excessive WPPA content may reduce the mechanical interlocking of the aggregate skeleton, as WPPA is more flexible and can deform under load and heat more than crushed stone. Correspondingly,

https://doi.org/10.31436/iiumej.v27i2.3843

the previous reasons make the asphalt pavement more susceptible to deformation and reduce its stability under heavy traffic loads, as excessive air voids and poor interlock between aggregates compromise its structural integrity [27, 31]. This finding is aligned with the results of a previous investigation [18]. Also, WPP (plastic) has a higher stiffness than the asphalt binder at 60 °C, the temperature of the Marshall test. The stiffnesses of WPP and the asphalt binder were determined from their softening points. Consequently, higher compression strength values were achieved for WPPA-modified samples [29].

In addition, the highest stability value was achieved for the modified mixture with fine WPPA at a replacement content of 2%, an increase of 34.8% compared to the control mixture. This is because the small WPPA coats the aggregate more readily, functioning similarly to the binder, while the coarse WPPA is attached to the aggregate particles, acting similarly to the natural aggregate [19, 32]. When mixed at high temperatures, the small particles of WPPA (less than 4.75 mm) can effectively coat the natural aggregates by forming a comprehensive and uniform thin film of softened or melted WPP around the aggregates, mainly because of their increased surface area-to-volume ratio, which improves the chemical and physical interactions with the aggregate particles as well as enhances overall binding [33].

Figure 8 shows that flow decreases as WPPA content increases across all modified mixtures, resulting in a less deformable, more rigid mixture. This is explained by polypropylene's lower specific gravity than pure asphalt, which allows particles to interpenetrate and improves aggregate interlock, increasing stability and reducing flow. Asphalt mixtures with moderate flow and high stability can significantly resist rutting distress and withstand heavy traffic loads. This result was aligned with the findings in the previous investigation [17]. However, the results for all modified mixtures met the minimum stability limit (minimum 8 kN) and the flow limit (allowed range 2-4 mm) specified in the local criteria [26].

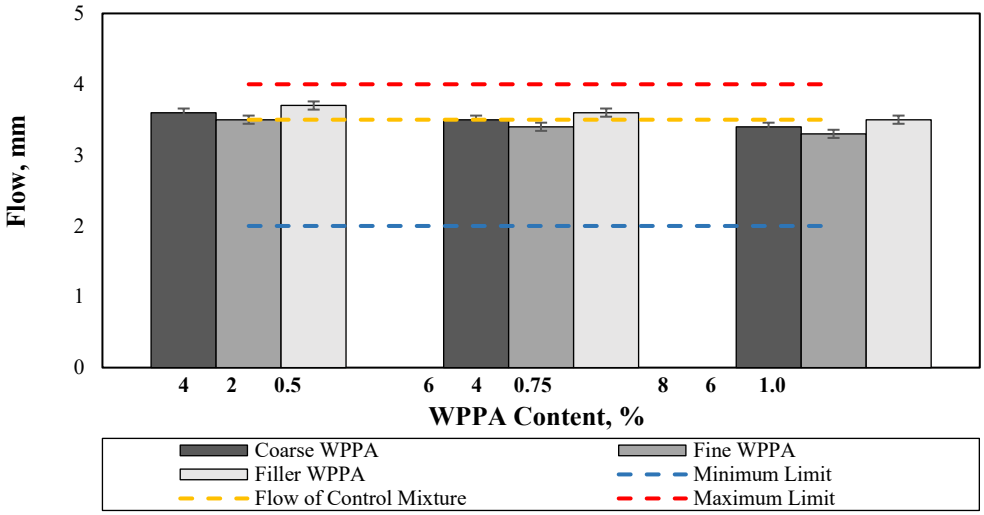


Figure 8. Marshall flow of control and WPPA-modified mixtures.

Furthermore, the Marshall test measures the stiffness or rigidity index of an asphalt mixture. The rigidity index is a ratio of stability to flow that evaluates the mixture's rut-proneness. In other words, it refers to the asphalt mixture's resistance to deformation under traffic loads. Higher rigidity index values, which indicate greater stiffness, enhance the mixture's ability to withstand prolonged deformation [34-35]. Table 5 shows the relationship

between WPPA concentration and rigidity index for both control and modified mixtures for filler, fine, and coarse WPPA. Mixtures become stiffer and rut-resistant when WPPA is added. The maximum rigidity index value for the WPPA-modified combination with 2% fine WPPA was 4.14 kN/mm. It was nearly 43% higher than that of the control mixture. The increase in the rigidity index is attributed to the distinctive impact of the rough surface texture and angular shape, which promote better mechanical interlock between aggregate particles and improve adhesion within the asphalt mixture's structure.

Table 5. Rigid index of control and WPPA-modified mixtures

Mixture Type	Control	WPPA-Modified								
		Coarse Replacement			Fine Replacement			Mineral Filler Replacement		
		WPPA Content, %								
		4	6	8	2	4	6	0.50	0.75	1.00
Rigid Index, kN/mm	2.90	3.53	3.29	2.95	4.14	3.44	2.93	3.68	3.89	3.53

4.2.2. Volumetric Properties

The bulk density (unit weight) values declined with the addition of WPPA in coarse, fine, and filler size replacements, as shown in Fig. 9. The density of the mixture decreases with increasing the content. For example, the density of the modified mixtures decreased by 3.7%, 7.5%, and 10.2% for coarse WPPA content of 4%, 6%, and 8%, respectively, compared to the control mixture. This pattern emphasizes the lightweight properties of WPPA and their impact on the mixture's bulk density. This implies that the WPPA-modified mixture is lighter than the control mixture. In general, asphalt mixtures with lower bulk density have lower stability. Despite this, WPPA-asphalt mixtures provide higher stability than the control mixture because WPPA enhances the bond between the natural aggregates and the asphalt binder. This unique bond improves cohesion and increases the stiffness of the asphalt mixture. In addition, the reduction in density can likely reduce asphalt mixture transportation and placement costs [36].

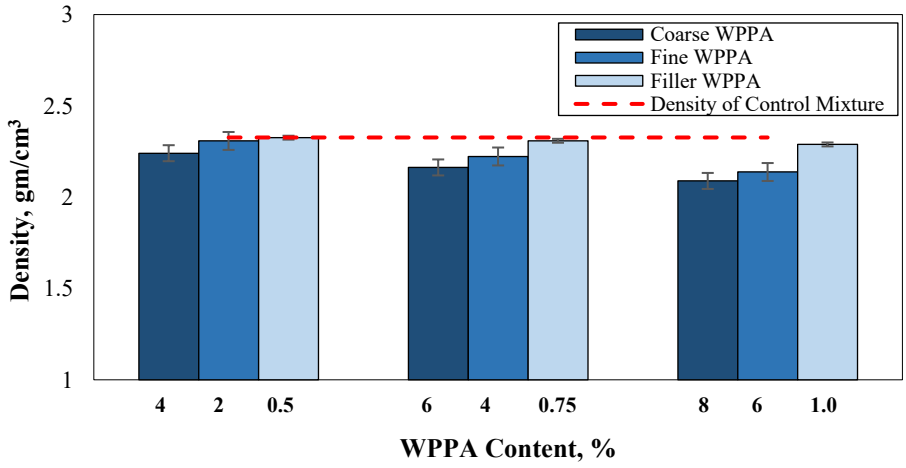


Figure 9. Bulk density of control and WPPA-modified mixtures.

Replacing natural aggregates with WPPA significantly affects the volumetric properties of the asphalt mixture, and a preliminary analysis of this effect was conducted. Table 6 provides the volumetric properties of control and WPPA-modified mixtures. The AV in the control mixture was almost 3.7%. Although there was an increase in AV across the modified mixtures,

it remained within the limited range of the surface layer (3–5%). This range was specified by the local road pavement specification to resist pavement distress such as fatigue cracking, bleeding, raveling, and water damage [37]. Obviously, the introduction of WPPA significantly increases the AV and VMA contents of asphalt mixtures of all sizes that were replaced, because, for the same weight, the low-density WPPA occupies a larger volume than natural aggregates, thereby affecting the overall volumetric properties of the asphalt mixtures. Overall, when WPPA enhances certain characteristics, excessive amounts compromise structural integrity.

Table 6. Volumetric properties of control and modified mixtures.

Mixture Type		Control	WPPA-Modified								
			Coarse Replacement			Fine Replacement			Mineral Filler Replacement		
			WPPA Content, %								
			4	6	8	2	4	6	0.50	0.75	1.00
Volumetri	AV, %	3.7	3.8	4.1	4.3	3.9	4.1	4.3	4.1	4.2	4.4
	VMA, %	14.2	14.25	14.31	14.35	14.75	14.81	15.09	15.85	15.98	16.01
	VFA, %	74	73	71	70	73	72	71	74	73	72

4.3. Indirect Tensile Strength

The impact of WPPA on the ITS for both control and modified asphalt mixtures is shown in Fig. 10. Based on the results, wet-conditioned samples exhibited lower values of ITS than dry-conditioned samples, reflecting the expected impact of moisture on the mixture's strength. Also, both dry and wet ITS readings were considerably enhanced by the addition of WPPA. This is attributed to the fact that the polypropylene polymer demonstrated higher resistance to aging and tensile degradation, making it the most appropriate option for high-performance applications, particularly for maintaining high moisture resistance. In addition, stronger bonding between the WPP polymer and aggregate prevents water from causing further damage in the asphalt mixture and enhances mixture strength in wet conditions, as the resulting bonding is hydrophobic. To clarify, the WPP is hydrophobic due to its non-polar molecular structure, which helps repel water and improve the asphalt mixture's resistance to moisture damage [38]. The WPPA also fills the microscopic voids between natural aggregate particles, thereby reducing pathways for water to enter the asphalt mixture structure and promoting better mechanical interlocking [19]. This bonding between WPP and aggregate creates a more cohesive, stronger matrix that enhances adhesion and resists water stripping. Accordingly, the pavement would be less susceptible to moisture damage. In conclusion, modified mixtures with 6% fine WPPA and 4% coarse WPPA exhibited the greatest dry ITS (1423 kPa) and wet ITS (1303 kPa), respectively.

The TSR results for the control and modified mixtures are presented in Fig. 11. TSR values, which range from 81.42% to 92.28% in the modified mixtures, indicate less susceptibility to moisture than in the control mixture. This improvement is likely due to the strong adhesion between the asphalt binder and WPPA, thereby reducing the moisture-induced stripping effect. In addition, the WPPA forms a three-dimensional network within the asphalt binder and creates a hydrophobic coating on the aggregate. Eventually, this strengthens the bond between the asphalt binder and aggregate particles, thereby improving the resistance of asphalt mixtures to moisture damage [39].

https://doi.org/10.31436/iiumej.v27i2.3843

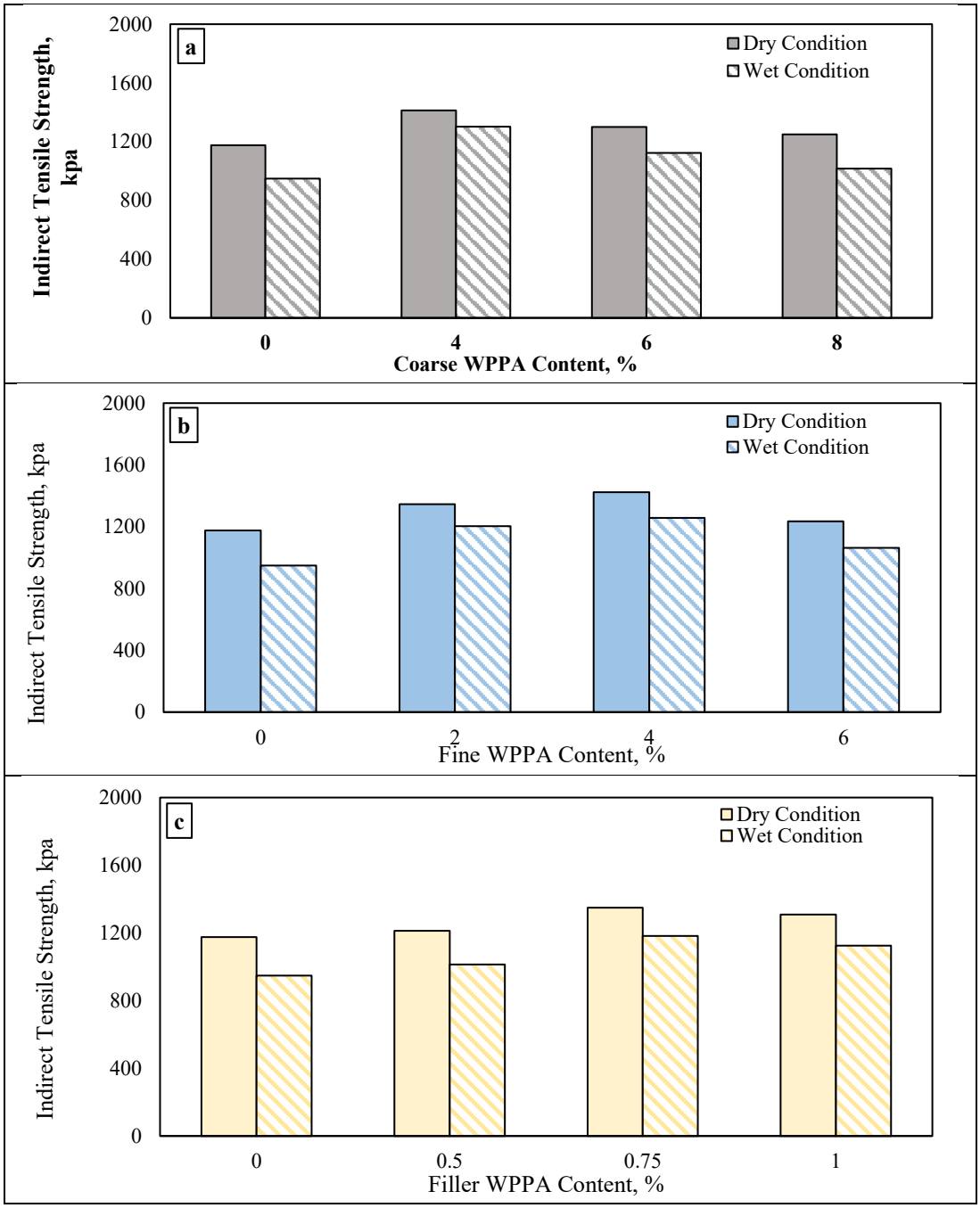


Figure 10. ITS results at dry and wet conditions: (a) coarse WPPA; (b) fine WPPA; and (c) filler WPPA.

ITS and TSR outcomes are consistent with the findings of the earlier work conducted by Lee et al. [19]. It is clear that when the WPPA content for all sizes increased, the TSR values decreased. Nevertheless, this finding does not coincide with previous research [29], which indicates a reduction in the TSR value with the addition of WPPA. This is due to the much stronger affinity between the asphalt binder and the natural aggregate than between the asphalt binder and waste plastic. Also, WPP in flake shape can negatively affect the water damage resistance of asphalt pavement because flake-shaped particles may reduce load-carrying capacity and increase voids. The acceptable TSR limit specified by Superpave requirements must be at least 80% for the wearing course. All modified mixtures fulfilled that requirement; even the control mixture had a TSR of 80.64%.

https://doi.org/10.31436/iiumej.v27i2.3843

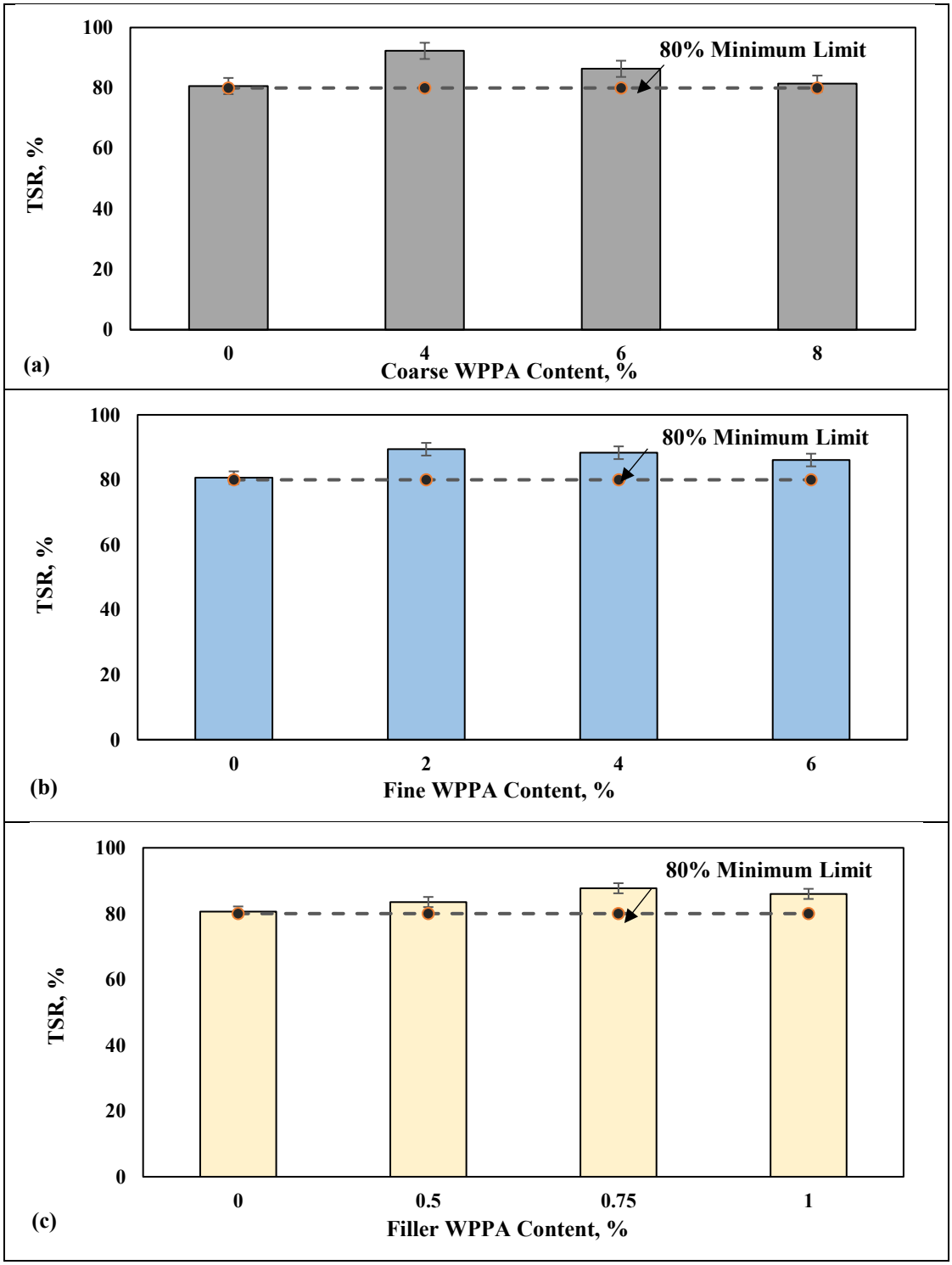


Figure 11. TSR of WPPA-modified mixtures: (a) coarse WPPA; (b) fine WPPA; and filler WPPA.

5. CONCLUSION

This study examined the influence of waste polypropylene plastic aggregate (WPPA) on the properties of asphalt mixtures and found that, when used at appropriate replacement levels, WPPA can improve abrasion resistance, Marshall stability, moisture resistance, and overall pavement performance. The hardness and resistance to tearing and wear increased with higher

coarse WPPA contents, showing improvements of 2.3%, 7.2%, and 10.2% at 4%, 6%, and 8% coarse WPPA replacement, respectively, compared with the natural aggregate mixture; this improvement is attributed to the lower surface roughness of WPPA, which reduces surface abrasion. However, the incorporation of WPPA also reduced the bulk density of the modified mixtures because natural aggregates with higher unit weight were partially replaced by lighter WPPA, resulting in a lighter aggregate skeleton. As the WPPA content increased, voids filled with asphalt (VFA) decreased, while air voids (AV) and voids in mineral aggregate (VMA) increased, mainly due to the lower density of WPPA and its tendency to rise or separate from heavier aggregate particles during mixing, placing, and compaction, particularly when excessive dosages or inadequate compaction are involved. In terms of mechanical performance, mixtures modified with coarse and fine WPPA generally showed improved Marshall stability, although higher replacement levels reduced stability by 6.5% for 8% coarse WPPA and 9.8% for 6% fine WPPA compared with the control mixture. In contrast, all mixtures containing WPPA as a filler material below 0.075 mm achieved higher stability than the control, likely because the fine WPPA particles formed a uniform, thin, molten film around the aggregates, thereby improving binder–aggregate interaction. The tensile strength ratio (TSR) also improved by approximately 14%, 11%, and 9% for mixtures containing 4% coarse WPPA, 2% fine WPPA, and 0.75% filler WPPA, respectively, and all WPPA-modified mixtures exceeded the 80% TSR threshold, confirming enhanced resistance to moisture-induced damage through stronger physical and chemical bonding between the binder and aggregate. The angular shape of WPPA further contributed to improved load-carrying capacity, mechanical interlocking, aggregate skeleton bonding, and adhesion within the asphalt mixture, suggesting the potential for longer pavement service life. Overall, the optimal WPPA replacement contents were identified as 4%, 2%, and 0.75% for coarse, fine, and filler sizes, respectively, because these proportions produced the best balance between strength, stiffness, stability, and moisture resistance, while higher dosages negatively affected mixture performance. Compared with the wet process, the dry-process incorporation of WPPA as an aggregate replacement may also more effectively support sustainability goals by enabling greater amounts of waste polypropylene plastic to be reused and diverted from landfills. Nevertheless, the study has three main limitations: high WPPA contents may reduce mixture homogeneity because WPPA, especially in fine sizes, tends to cluster with aggregate particles; producing WPPA with controlled sizes and dimensions may be more costly than conventional cutting and shredding methods; and the use of WPP collected from different sources may introduce contamination and material variability, making quality control difficult and potentially causing unpredictable pavement performance when WPPA is incorporated into asphalt mixtures.

6. RECOMMENDATIONS

Based on the findings of this research, several recommendations are proposed to strengthen the future development and practical application of waste polypropylene plastic (WPP) in asphalt mixtures. Future studies should investigate alternative approaches for incorporating WPP, including modified mixing techniques that can improve mixture performance while controlling production cost and environmental impact. For example, the use of a semi-wet process to modify asphalt binders with WPP may offer a practical alternative to conventional dry mixing by reducing emissions and improving binder–polymer interaction. Since the present findings are mainly derived from laboratory experiments, field trials are strongly recommended to validate the long-term performance of WPPA-modified asphalt pavements under actual traffic loading, weather exposure, and construction conditions. Further research should also include additional mechanical performance tests, particularly fatigue resistance and rutting evaluation, to better understand the durability of asphalt mixtures incorporating WPPA

produced through the recycling process adopted in this study. In addition to technical performance, future work should examine the environmental implications of using WPP in asphalt mixtures through comprehensive assessments such as life cycle assessment, with specific attention to carbon emissions, energy consumption, and potential microplastic release. Such evaluation is necessary to provide a more balanced understanding of the sustainability benefits and possible environmental risks associated with WPP utilization in pavement materials. Finally, advanced chemical and physical characterization of WPP should be conducted before its incorporation into asphalt mixtures to ensure material consistency, quality control, and predictable pavement performance.

REFERENCES

[1] Kibria MG, Masuk NI, Safayet R, Nguyen HQ, Mourshed M. (2023) Plastic waste: challenges and opportunities to mitigate pollution and effective management. Int J Environ Res., 17(1): 20. https://doi.org/10.1007/s41742-023-00507-z

[2] Organization for Economic Co-operation and Development (OECD). (2022) Plastic pollution is growing relentlessly as waste management and recycling fall short, says OECD. [www.oecd.org/en/about/news/press-releases/2022/02/plastic-pollution-is-growing-relentlessly-as-waste-management-and-recycling-fall-short.htm]

[3] Lingyun Y, Zhengwu L, Zhanping Y, Dongdong G, Xu Y, Fu X, Mohammad H, Aboelkasim D. (2022) Review of recycling waste plastics in asphalt paving materials. J Traffic Transp Eng (Engl Ed), 9(5): 742-764. https://doi.org/10.1016/j.jtte.2022.07.002

[4] Basheet SH, Latief RH. (2024) Asphalt binder modification with high-density polyethylene polymer and low-density polyethylene polymer–efficiency of conducting semi-wet mixing process. J Ecol Eng., 25(12): 202-212. https://doi.org/10.12911/22998993/194397

[5] Yuetan M, Hongyu Z, Xi J, Pawel P, Rui X, Miaomiao Z, Baoshan H. (2021) The utilization of waste plastics in asphalt pavements: A review. Clean Mater., 2: 100031. https://doi.org/10.1016/j.clema.2021.100031

[6] Soheil H, Ailar H, Nioushasadat HSJ, Nasser K. (2021) The use of plastic waste in asphalt: a critical review on asphalt mix design and Marshall properties. Constr Build Mater., 309: 125185. https://doi.org/10.1016/j.conbuildmat.2021.125185

[7] Basheet SH, Latief RH. (2024) Assessment of the properties of asphalt mixtures modified with LDPE and HDPE polymers. Int J Appl Sci Eng., 21(5):1-11. https://doi.org/10.6703/IJASE.202412_21(5).007

[8] John EE, Paula H, Tim AO. (2025) Processing behavior evolution of recycled polypropylene: An integrated experimental and Computer-Aided engineering simulation study. Phys. Fluids, 37: 033110. https://doi.org/10.1063/5.0260486

[9] Hossain MT, Shahid MA, Mahmud N, Habib A, Rana MM, Khan SA, Hossain MD. (2024) Research and application of polypropylene: a review. Discov Nano, 19(1): 2. doi: 10.1186/s11671-023-03952-z

[10] Ameur AB, Valentin J, Baldo N. (2025) A review on the use of plastic waste as a modifier of asphalt mixtures for road constructions. CivilEng, 6(2): 17. https://doi.org/10.3390/civileng6020017

[11] Guo F, Pei J, Zhang J, Xue B, Sun G, Li R. (2020) Study on the adhesion property between asphalt binder and aggregate: A state-of-the-art review. Constr Build Mater., 256: 119474. https://doi.org/10.1016/j.conbuildmat.2020.119474.

[12] Xiang H, Wang Z, Deng M, Tan S, Liang H. (2025) Adhesion characteristics of an asphalt binder-aggregate interface based on molecular dynamics. Materials (Basel), 18(5): 981. https://doi.org/10.3390/ma18050981

[13] Hasan U, Whyte A, Al Jassmi H, Hasan A. (2022) Lifecycle cost analysis of recycled asphalt pavements: determining cost of recycled materials for an urban highway section. CivilEng, 3(2): 316-331. https://doi.org/10.3390/civileng3020019

[14] NCAT; WRI; GHK; Dow. (2021) Performance properties of laboratory produced recycled plastic modified (rpm) asphalt binders and mixtures. NCHRP Project 9-66 Interim Report; National Center of Asphalt Technology: Auburn, AL, USA.

[15] Hall F, White G. (2021) The effect of waste plastics on the ageing phenomenon of bituminous binders and asphalt mixtures. Mater., 14(20): 6176. https://doi.org/10.3390/ma14206176

[16] Xu F, Zhao Y, Li K. (2022) Using waste plastics as asphalt modifier: a review. Mater., 15(1): 110. https://doi.org/10.3390/ma15010110

[17] Kim Y, Kim K. (2025) Evaluation of thermal aging susceptibility of recycled waste plastic aggregates (low-density polyethylene, high-density polyethylene, and polypropylene) in recycled asphalt pavement mixtures. Polymers, 17: 731. https:// doi.org/10.3390/polym17060731

[18] Daniela LB, Puiu LG, Gabriel BC, Viorica G. (2023) Recycling micro polypropylene in modified hot asphalt mixture, Sci Rep., 13: 3639, https://doi.org/10.1038/s41598-023-30857-9

[19] Lee SY, Kim KW, Yun Y, Le THM. (2024) Evaluation of eco-friendly asphalt mixtures incorporating waste plastic aggregates and additives: Magnesium, fly ash, and steel slag. Case Stud Constr Mater., 20: e02756. https://doi.org/10.1016/j.cscm.2023.e02756

[20] Dione A, Faye P, Thiam M, Diouf B, Diallo C, Becker A, Diouf D, Sow M, Thiam M, Ba M. (2023) Evaluation of the influence of plastic waste in improving the mechanical characteristics of asphalt mixes (Senegal, West Africa). Open J Civ Eng., 13: 517-527. doi: 10.4236/ojce.2023.133037

[21] Li Y, Zhao C, Li R, Zhang H, He Y, Pei J, Lyu L. (2025) Dry-process reusing the waste tire rubber and plastic in asphalt: modification mechanism and mechanical properties. Constr Build Mater., 458: 139759. https://doi.org/10.1016/j.conbuildmat.2024.139759

[22] Movilla-Quesada D, Raposeiras AC, Olavarría J. (2019) Effects of recycled polyethylene terephthalate (PET) on stiffness of hot asphalt mixtures. Adv Civ Eng., 2019: 6969826. https://doi.org/10.1155/2019/6969826

[23] Aldagari S, Kabir, SF, Fini EH. (2021) Investigating aging properties of bitumen modified with polyethylene-terephthalate waste plastic. Resour Conserv Recycl., 173: 105687. https://doi.org/10.1016/j.resconrec.2021.105687

[24] Haider S, Hafeez I, Jamal, Ullah R. (2020) Sustainable use of waste plastic modifiers to strengthen the adhesion properties of asphalt mixtures. Constr Build Mater., 235: 117496. https://doi.org/10.1016/j.conbuildmat.2019.117496

[25] Ramli I, Yaacob H, Abdul Hassan N, Ismail CR, Hainin MR. (2015) Fine aggregate angularity effects on rutting resistance of asphalt mixture. J Teknol., 65(3): 105-109. https://doi.org/10.11113/jt.v65.2154

[26] State Corporation for Roads and Bridges. General Specification for Roads and Bridges (SORB/R9)-Hot Mix Asphaltic Concrete Pavement. Baghdad, Iraq: Department of Planning and Studies, Republic of Iraq, Ministry of Housing and Construction; 2003.

[27] Franesqui MA, Rodríguez-Alloza AM, García-González C. (2023) Reuse of plastic waste in asphalt mixtures with residual porous aggregates. Case Stud Constr Mater., 19: e02361. https://doi.org/10.1016/j.cscm.2023.e02361

[28] Sánchez DB, Airey G, Caro S, Grenfell J. (2020). Effect of foaming technique and mixing temperature on the rheological characteristics of fine RAP-foamed bitumen mixtures. Road Mater Pavement Des., 21(8): 2143–2159. https://doi.org/10.1080/14680629.2019.1593228

[29] Movilla-Quesadaa D, Raposeirasa AC, Silva-Kleina, LT, Lastra-Gonzálezb P, Castro-Fresno, D. (2019) Use of plastic scrap in asphalt mixtures added by dry method as a partial substitute for bitumen. Waste Manag., 87: 751–760. https://doi.org/10.1016/j.wasman.2019.03.018

[30] Basheet SH, Latief RH. (2025) The impact of using polyethylene polymer on the properties of hot asphalt mixture by conducting semi-wet and dry mixing process. Int J Eng., Trans B, Appl., 38(05): 1108-1119. https://doi.org/10.5829/ije.2025.38.05b.13

[31] Bueno IM, Teixeira JESL. (2024) Waste plastic in asphalt mixtures via the dry method: a bibliometric analysis. Sustain., 16: 4675. https://doi.org/10.3390/su16114675

[32] Assefa N. (2021) Evaluation of the effect of recycled waste plastic bags on mechanical properties of hot mix asphalt mixtures for road construction. Sustain Environ., 7(1): 1957649. https://doi.org/10.1080/27658511.2021.1957649

[33] Boom YJ, Xuan DL, Enfrin M, Swaney M, Masood H, Pramanik BK, Robert D, Giustozzi F. (2023) Engineering properties, microplastics and emissions assessment of recycled plastic modified asphalt mixtures. Sci Total Environ., 893: 164869. https://doi.org/10.1016/j.scitotenv.2023.164869

[34] Mohsin HK, Latief RH. (2025) Natural bitumen in hot asphalt mixture: suitability of using treated natural bitumen instead of petroleum asphalt binder. IIUM Eng J., 26(2): 27-50. https://doi.org/10.31436/iiumej.v26i2.3452

[35] Weiguang Z, Adnan K, Ju H, Jingtao Z, Tianyi P, Hanglin C. (2021) Predicting Marshall parameters of flexible pavement using support vector machine and genetic programming. Constr Build Mater., 306: 124924. https://doi.org/10.1016/j.conbuildmat.2021.124924

[36] Josué C, Arminda A, João S, Adelino F. (2025) Incorporation of plastic waste into road pavements: Performance of SMA mixtures containing flakes of low-density polyethylene. Constr Build Mater., 471: 140766. https://doi.org/10.1016/j.conbuildmat.2025.140766

[37] Mohsin HK, Latief RH. (2025) Properties evaluation of natural bitumen-filler mastic mixture. Eng App Sci Res., 52(1): 66-80. doi: 10.14456/easr.2025

[38] Yang N, Du C, Tang Y, Li Z, Xu S, Xu X. (2025) Waste polypropylene in asphalt pavements: a state-of-the-art review toward circular economy. Sustain., 17(24): 10954. https://doi.org/10.3390/su172410954

[39] Mturi G, Ncolosi N, O'Connell J, Simelane M. (2025) The incorporation of the plastic-coated aggregates into a South African asphalt mixture. Road Mater Pavement Des., 26(10): 2633–2648. https://doi.org/10.1080/14680629.2025.2460483

https://doi.org/10.31436/iiumej.v27i2.3858

Enhancement of Stabilization/Solidification Method with Bioaugmentation for Sludge Remediation

MUHAMMAD HASANUDDIN MOHD KAMAL[1], NASUHA MOHD RAIS[1], NOR AMANI FILZAH MOHD KAMIL[1,2*], NUR FATEHA MD NOR[3]

[1]*Faculty of Civil Engineering and Built Environment, University Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia*
[2]*Micropollutant Research Centre, University Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia*
[3]*See Sen Chemical Berhad, Sendayan Techvalley Industrial Estate, 71950 Seremban, Negeri Sembilan, Malaysia*

**Corresponding author: noramani@uthm.edu.my*

(Received: 27 July 2025; Accepted: 2 March 2026; Published online: 10 May 2026)

ABSTRACT: The Stabilization/Solidification (S/S) method encapsulates heavy metals in sludge, while bioaugmentation can degrade organic contaminants. However, each method is limited in its ability to address both contaminants simultaneously. This study aims to enhance the S/S method by integrating bioaugmentation, using either immediate or pre-inoculation of bio-agents. Ordinary Portland Cement (OPC) served as the binder in the S/S process, with the bacteria *Sphingobacterium spiritivorum* and the fungi *Aspergillus brasiliensis* acting as bio-agents. Before testing, chemical oxygen demand (COD) and heavy metals in sludge and cement were measured. COD and heavy metal concentrations' leaching behavior were tested using TCLP after 14 and 28 days. The result shows 23.39% of organic contaminants were in the fibreboard sludge. The TOC results for Portland cement were negligible. All heavy metal elements (Cr, Cu, Zn, As, and Pb) in the Portland cement and fibreboard sludge exceeded the standard USEPA limit. COD reduction was evaluated for immediate versus pre-inoculation of fungi and bacteria over 14 and 28 days. Pre-inoculation consistently yielded lower COD levels than immediate inoculation, with fungi pre-inoculation achieving the greatest reduction (579 mg/L at 14 days; 544.7 mg/L at 28 days) and bacteria pre-inoculation showing moderate improvement (579.3 mg/L at 14 days; 570.3 mg/L at 28 days). Concentrations of heavy metal reductions followed a similar trend: after 28 days, pre-inoculated samples contained 0.01 mg/L (Cr), 0.22 mg/L (Cu), 0.06 mg/L (Zn), and 0 mg/L for both As and Pb, whereas immediate-inoculated samples showed 0.04 mg/L (Cr), 0.22 mg/L (Cu), 0.01 mg/L (Zn), and 0 mg/L for As and Pb. Statistical comparison shows significant differences between immediate-inoculation and pre-inoculation, but not between the bio-agents used in this study. To improve the findings, a co-culture method that combines both bio-agents is recommended, offering a viable strategy for industrial sludge management, with future studies needed to scale up the method for practical, sustainable waste treatment applications.

ABSTRAK: Kaedah Pengukuhan dan Penstabilan (S/S) bertujuan mengkapsulkan logam berat dalam enap cemar, manakala bioaugmentasi digunakan bagi mengurai bahan pencemar organik. Walau bagaimanapun, setiap kaedah mempunyai keterbatasan dalam menangani kedua-dua jenis pencemar secara serentak. Kajian ini bertujuan menambah baik kaedah S/S dengan menggabungkan bioaugmentasi, sama ada melalui inokulasi segera atau pra-inokulasi mikrob. Simen Portland Biasa digunakan sebagai agen pengikat dalam proses S/S, manakala bakteria *Sphingobacterium spiritivorum* dan kulat *Aspergillus brasiliensis* bertindak sebagai agen bio. Sebelum kajian dijalankan, Permintaan Oksigen Kimia (COD) dan logam berat dalam enap cemar serta simen ditentukan. Tingkah laku larut lesap COD dan kepekatan logam

https://doi.org/10.31436/iiumej.v27i2.3858

berat diuji menggunakan kaedah TCLP selepas 14 dan 28 hari. Dapatan kajian menunjukkan sebanyak 23.39% bahan pencemar organik terdapat dalam enap cemar papan gentian. Kandungan TOC dalam simen Portland didapati sangat rendah dan tidak memberi kesan ketara. Semua unsur logam berat (Cr, Cu, Zn, As dan Pb) dalam Simen Portland Biasa dan enap cemar papan gentian melebihi had piawaian yang ditetapkan oleh USEPA. Pengurangan COD dinilai bagi inokulasi segera berbanding prainokulasi dengan kulat dan bakteria selama 14 dan 28 hari. Prainokulasi secara konsisten menghasilkan tahap COD lebih rendah berbanding inokulasi segera, dan prainokulasi kulat menunjukkan pengurangan paling ketara (579 mg/L pada 14 hari; 544.7 mg/L pada 28 hari) manakala prainokulasi bakteria menunjukkan peningkatan sederhana (579.3 mg/L pada 14 hari; 570.3 mg/L pada 28 hari). Kepekatan pengurangan logam berat mengikuti corak serupa: selepas 28 hari, sampel prainokulasi mengandungi 0.01 mg/L (Cr), 0.22 mg/L (Cu), 0.06 mg/L (Zn), dan 0 mg/L bagi As dan Pb, manakala sampel inokulasi segera menunjukkan 0.04 mg/L (Cr), 0.22 mg/L (Cu), 0.01 mg/L (Zn), dan 0 mg/L bagi As dan Pb. Perbandingan statistik menunjukkan terdapat perbezaan signifikan antara kaedah inokulasi segera dan prainokulasi, namun tiada perbezaan signifikan antara jenis agen bio yang digunakan dalam kajian ini. Bagi meningkatkan keberkesanan rawatan, kaedah ko-kultur menggabungkan kedua-dua bioagen dicadangkan, menawarkan potensi terbaik bagi pengurusan sisa industri. Kajian lanjut diperlukan untuk penskalaan bagi aplikasi praktikal dan rawatan sisa lestari.

KEYWORDS: *Biodegradation, Hazardous Waste, Emerging Pollutant, Remediation*

1. INTRODUCTION

The increasing number of factories in Malaysia has led to an increase in sludge production in industrial wastewater. The amount of waste produced by industrial activities is also variable due to landfill overcapacity and other waste sources. Demand for forest products, including timber for construction, paper, and furniture, has also increased, which corresponds to the rapid population and urbanization growth. In this research, sludge production focuses only on sludge from industrial wastewater treatment plants used in fibreboard production. Fibreboard is a product made from rubber and tropical trees. This scenario has led to an increase in sludge production from wastewater treatment.

The sludge produced in the wastewater treatment facility was obtained from the fibreboard cleaning procedure. The wastewater process completed several steps before the sludge formation. About 500kg of fibreboard sludge was produced by Evergreen Fibreboard Berhad in a month, and the licensed contractor's transportation cost was approximately RM500 per month. After the sludge was delivered to Kualiti Alam Sdn Bhd, a specific procedure was carried out before it was disposed of at the landfill [1]. The Resources Conservation and Recovery Act (RCRA) states that the sludge was categorized as scheduled waste, also known as hazardous waste, requiring specific treatment for safe disposal.

Sludge production contains high concentrations of heavy metals, organic contaminants, and pathogenic microorganisms. A detrimental effect will occur if the sludge is not properly treated. Thus, sludge produced by fibreboard processing may threaten the surrounding environment. The factory is located near the university campus, residential zones, and other developed infrastructure, potentially posing health risks to nearby communities, including respiratory infections and skin problems. As a result, a polluted environment has disturbed the sustainability of life in that area. Besides, sludge production has affected agricultural activities and water quality in nearby locations [2].

The Stabilization/Solidification (S/S) method was used to immobilize the heavy metals. Based on previous studies, the S/S method was less efficient for organic contaminants when

using only cement as the binder [3]. S/S alone resulted in higher COD concentrations, whereas bioaugmentation achieved the lowest COD levels [13]. Moreover, COD levels in the S/S method-only samples increased between 14 and 28 days, indicating limited degradation of organic matter, whereas bioaugmented samples exhibited significant COD reductions over the same period [33]. A high concentration of calcium oxide (CaO) in the cement has made it compatible with heavy metals but not with organic contaminants [4]. The S/S method was less suitable for organic waste due to organic contaminants that prevented binder hydration and typically did not form chemical bonds in binder hydration solutions [5]. A key limitation of the S/S method is its inability to eliminate contaminants, as they remain within the solidified matrix. This has seriously threatened the safety of solidified samples when the binder did not adsorb or react with pollutants, producing an unstable component [6].

Therefore, bioaugmentation was incorporated into the S/S process to mitigate the adverse effects of sludge on human health and the environment and to enhance the removal of organic contaminants. The bioaugmentation degraded the organic contaminants into safe intermediates and end products. The addition of bioaugmentation to the S/S method was inspired by the study conducted by [3], which reported that a bacterial consortium and Portland cement were employed to evaluate the potential for Total Petroleum Hydrocarbons (TPH) degradation within granulated drill cuttings stabilized/solidified through bacterial integration. Then, a compost composed of bacteria and nutrient sources was added to S/S matrices to reduce hydrocarbons in petroleum drill cuttings before disposal [7]. Previous studies have demonstrated that the combination method can degrade both heavy metals and organic contaminants. Thus, the combination of S/S and the bioaugmentation method was conducted in this study to enhance the degrading ability towards heavy metals and organic contaminants.

The bioaugmentation treatment method was shown to have high potential for degrading organic contaminants. Besides, contaminated matrices can be converted into stable, reusable products through biological treatment [8]. However, both groups of contaminants are present in the sludge at high concentrations. Thus, combining both methods is required to ensure that organic contaminants are degraded. This combination treatment method must be integrated to eliminate contaminants over time [7]. Bioaugmentation is a technology that improves biomass by introducing specialized microorganisms into the natural flora. The bioaugmentation method may enhance bioremediation efficiency [9]. Chemical Oxygen Demand (COD) and Total Nitrogen (TN) are commonly degraded during bioaugmentation treatment, and microbes are commonly used as bioagents.

However, previous studies have only demonstrated the effectiveness of the combination method when adding bacteria as bio-agents. Research on the addition of fungi to cementitious systems is limited. In this study, the combination of S/S and bioaugmentation was conducted by inoculating the S/S matrices with microbes (Sphingobacterium spiritivorum) and fungi (Aspergillus brasiliensis) as bio-agents. The bacteria were chosen in this study because they were used in the study by [10] to degrade organic contaminants in soil and had been shown to improve the degradation of organic contaminants by inoculating Sphingobacterium spiritivorum strains in the soil field. Then, *Aspergillus brasiliensis* was chosen as a bioagent in this study due to its ability to produce organic acids effective in metal bioleaching, such as tartaric, oxalic, citric, and malic [11]. In a bioremediation treatment, *Aspergillus sp.* identified from tannery waste was found to be highly effective in reducing the Chemical Oxygen Demand (COD) and color of tannery effluent [12]. In this study, the results between these two microbes were compared.

https://doi.org/10.31436/iiumej.v27i2.3858

2. METHODOLOGY

2.1. Raw Materials Preparation

This study utilized fibreboard sludge and Portland cement as its primary materials. Approximately 40 kg of sludge was acquired from a fiberboard manufacturing factory prior to the sanitation stage of its wastewater treatment process. Before testing, the sludge was kept at room temperature in a sealed plastic bag to ensure its viability. Meanwhile, Portland cement was procured from a nearby hardware store as the binding agent.

2.1.1. Bacteria Preparation

The bacteria (*Sphingobacterium spiritivorum*) used in this study were isolated from sewage sludge and stored in microbeads (microbankTM) at -80 °C. 200 ml of distilled water and 5.6 g of Nutrient Agar (C1225) powder were combined in a bottle (USEPA, Method FNES6). Subsequently, the bottle was shaken 3 to 4 times and autoclaved for approximately 45 minutes to achieve sterilization. Next, to ensure the agar solidified, the agar mixture was transferred into several Petri dishes and allowed to stand for 10 to 12 minutes. An inoculation loop was used to acquire a single colony of *S. spiritivorum*. The inoculation loop was streaked on a Petri dish using the quadrant method, which reduces the bacterial load at each streak to separate individual bacterial cells. A biohazard safety cabinet was used for these procedures. Since the bacterial isolates were active on days 3 or 4, the petri dishes were incubated upside-down at 37ºC for 3 to 4 days, as shown in Figure 1. The parafilm closure was used. A temperature range of 30–37 °C was used as the optimal temperature for the bacteria.

A small number of frozen beads were transferred to universal bottles containing nutrient broth, and the broth was incubated at 30°C for 3 days to resuscitate the bacteria. In a bottle intended for a cube mould, 26 g of Nutrient Broth powder and 2 L of distilled water were combined to prepare bacterial broth (USEPA, Method FNES6). An autoclave machine was used to sterilize the broth for approximately 45 minutes. Four Petri dishes containing growing bacteria were combined with the sterilized broth, and the experiment was carried out in a biohazard safety cabinet. Before utilizing the broth as a bio-agent in the bioaugmentation, the bottle had to be incubated for 3 days at 30°C.

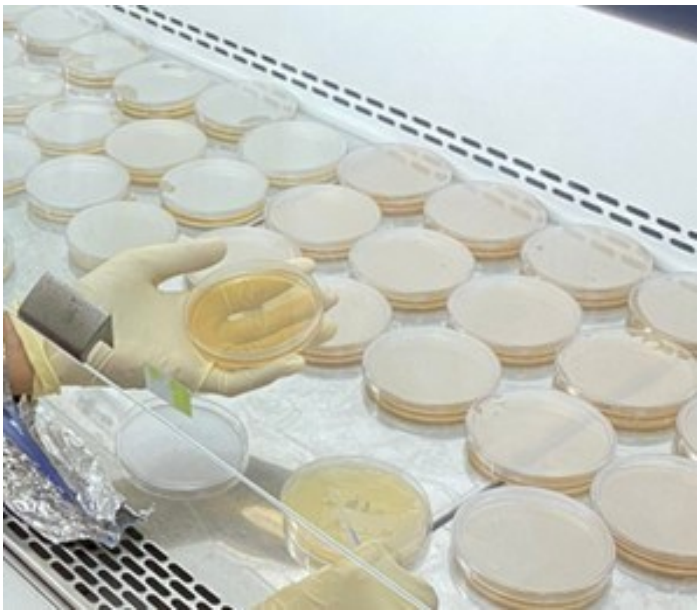


Figure 1. Process of bacterial culture in a petri dish.

https://doi.org/10.31436/iiumej.v27i2.3858

2.1.2. Fungi Preparation

The fungal strain used in this study was *Aspergillus brasiliensis* ATCC® 16404TM. Initially, 200 mL of distilled water and 7.8 g of potato dextrose agar (PDA) powder were combined in a bottle using the USEPA Method FNES6. The bottle was then shaken 3 to 4 times and autoclaved for approximately 45 minutes to sterilize it. Then, to ensure the agar solidified, the PDA agar mixture in the bottle was poured into several Petri dishes and allowed to sit for 10 to 12 minutes. After that, 1 ml of the fungal spore suspension was added to the PDA agar. The biohazard safety cabinet served as the place for these procedures. As sporulation was completed after 7 days, the Petri dishes were covered with parafilm and incubated at 28ºC for 7 days. The ideal temperature for the fungi was 30ºC.

The USEPA standard process, Method FNES6, was used to produce the fungi broth. First, 2 L of Ultra-Pure Water (UPW) and 48 g of potato dextrose broth powder were combined in a bottle (Figure 2). An autoclave was used to sterilize the broth for approximately 45 minutes. In a biohazard safety cabinet, 4 Petri dishes containing growing fungus were combined with the sterilized broth. Before the solution was added to the cement and sludge mixture as a bio-agent, the bottle was shaken 4 to 5 times.



Figure 2. Process of mixing fungi into Ultra-Pure Water (UPW)

2.2. Measurements of Organic Contaminants and Heavy Metals Raw Materials.

Total Organic Carbon (TOC) analysis was used to quantify the carbon content of the sludge sample's organic compounds. The sludge sample was prepared as a powder prior to analysis. The TOC analysis was performed in accordance with EPA Method 9060A.

This research used the Acid Digestion method (USEPA, Method 3050B) to determine the concentration of heavy metals in sludge. Acid digestion is a common preparation method that releases heavy metals into solution by dissolving complex sample matrices, making them easier to analyze. This process guarantees that the heavy metals are present in a form that can be precisely measured with a variety of analytical methods, including Inductively Coupled Plasma Mass Spectrometry (ICP-MS), Atomic Absorption Spectroscopy (AAS), and Inductively Coupled Plasma Optical Emission Spectrometry (ICP-OES).

https://doi.org/10.31436/iiumej.v27i2.3858

2.3. Design of S/S Method with Immediate-Inoculation and Pre-Inoculation of Bio-agent

This study focused on the hybrid method of immediate-inoculation and pre-inoculation. The immediate-inoculation and pre-inoculation were applied with bacteria (*Sphingobacterium spiritivorum*) and fungi (*Aspergillus brasiliensis*) as augmentation agents in the bioaugmentation method. For the immediate-inoculation sample, augmentation agents and cement were mixed simultaneously with Portland cement and cured for 14 days and 28 days to produce the S/S matrices. Meanwhile, for the pre-inoculation sample, augmentation agents were inoculated and stored at room temperature for 7 days before mixing with Portland cement and curing for the same period as the immediate-inoculation sample. Portland cement serves as the binder in the S/S method. For the control sample, no augmentation agents were inoculated in this sample. Only indigenous microorganisms were used for contaminant removal.

The optimum Portland cement-to-sludge ratio (70:30) employed in this study was chosen following [13]. Subsequently, 10×10^{10} CFU/ml bacterial and fungal strains were added to the hybrid method. Table 1 shows the preparations of bacteria and fungi for immediate inoculation and pre-inoculation.

Table 1. Preparation of bacteria and fungi inoculation for immediate inoculation and pre-inoculation

Sample	Portland Cement (%)	Sludge (%)	Sludge + Bacteria (%)	Sludge + Fungi (%)
14 Days				
Fungi pre-inoculation	70	-	-	30
Fungi's immediate inoculation	70	-	-	30
Bacteria pre-inoculation	70	-	30	-
Bacteria's immediate inoculation	70	-	30	-
Control sample	70	30	-	-
28 Days				
Fungi pre-inoculation	70	-	-	30
Fungi's immediate inoculation	70	-	-	30
Bacteria pre-inoculation	70	-	30	-
Bacteria's immediate inoculation	70	-	30	-
Control sample	70	30	-	-

The S/S matrices were prepared and cured for 14 days and 28 days using Portland cement as the binder. Gradually add water to the mixture, mixing until no lumps remain. The homogenized mixture was then cast into 100 mm³ molds in three layers, with each layer manually compacted using a tamping rod for at least 50 hits. In order to guarantee that air curing occurs at a regulated temperature and humidity condition, the samples were triplicated for two hydration durations, which were 14 and 28 days, following the casting process. The concrete cubes were subsequently subjected to compressive strength testing. Figure 3 shows the flowchart of the S/S matrices design.

https://doi.org/10.31436/iiumej.v27i2.3858

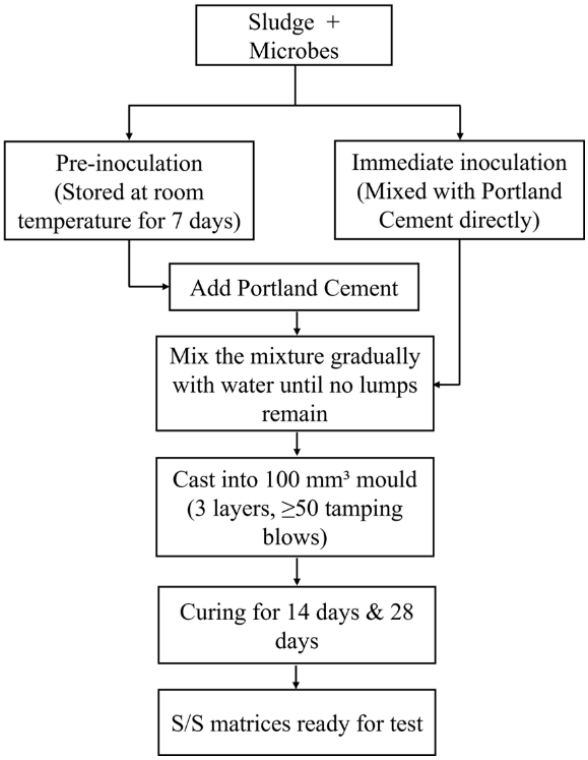


Figure 3. Flowchart of the S/S matrices design process

2.4. Compressive Strength Test

After curing, the concrete cubes were removed and thoroughly dried. Sample numbers corresponded to those listed in Table 1. Prior to testing, the dimensions and weight of each cube were recorded. Each cube was positioned vertically on the base plate of a compression testing machine, as shown in Figure 4(a), and the upper plate was lowered until it contacted the cube surface. The load was applied gradually, and the maximum compressive load at initial cracking or failure was recorded. Compressive strength was calculated by dividing this maximum load by the cube's cross-sectional area. This test is critical, as a strong S/S matrix physically immobilizes hazardous substances, whereas weak or cracked solidified materials may allow leaching of contaminants. Figure 4(b) illustrates the concrete cube after undergoing the compressive strength test, showing the point of failure.



(a) (b)

Figure 4. (a) Compressive strength test and (b) Concrete cube of S/S matrices samples

https://doi.org/10.31436/iiumej.v27i2.3858

2.5. Leachability Test

The Toxicity Characteristic Leachability Procedure (TCLP) test required that cube samples be crushed to a size of less than 9.5 mm to meet the USEPA standard (Method 1311). 5.7 ml of concentrated acetic acid, CH3COOH, and 1000 ml of distilled water were added to a conical flask, half-filled, and the mixture was shaken to mix. After that, the 1000 ml conical flask was filled with distilled water again to the top limit line and shaken again. The pH of 400 ml of reagent water, prepared by mixing 5.7 ml of concentrated CH3COOH with distilled water, was measured in a beaker. Subsequently, 20 g of crushed cube sample was added to the reagent water, and the mixture was placed in screw-capped polyethylene bottles. Then, the bottles were rotated at 30 rpm for 18 hours in a rotary chamber (Figure 5(a). The collected leachate was filtered using 7 µm fiber filters, as depicted in Figure 5(b). Concentrated nitric acid (HNO3) was used to store the leachate, and the pH was kept below 2.



(a) (b)

Figure 5. (a) Rotary chamber and (b) Filtration process



(a) (b)

Figure 6. (a) HACH DR600 equipment and (b) Inductively Coupled Plasma Mass Spectroscopy (ICP-MS)

https://doi.org/10.31436/iiumej.v27i2.3858

A Chemical Oxygen Demand (COD) test was used to indirectly quantify the total organic content of the collected leachate. This COD test also displayed the mass of oxygen utilized per liter of solution. The COD test also measured the amount of oxygen consumed by water pollutants during chemical oxidation. First, a 2 ml diluted, filtered sample was placed in High Range COD vials, and the leachate was kept in concentrated nitric acid (HNO3) at pH <2. It was heated in the DRB200 reactor for 2 hours. The concentration of organic contaminants in the leachate was subsequently determined using analytical methods, with measurements conducted on the DR6000 equipment, as illustrated in Figure 6(a). The test was carried out using the USEPA Method 410.3 and the EPA-600/4-79-020 standard procedure. Heavy metal concentration in the extracted and measured samples was then examined using an ICP-MS analyzer, as depicted in Figure 6(b).

3. RESULTS

3.1. Concentration of Organic Contaminants in Raw Materials

Analysis of Total Organic Carbon (TOC) offered important information on the organic content in the fibreboard sludge. The TOC analysis measured SSM-TC (Total Carbon), SSM-IC (Inorganic Carbon), and SSM-TOC (Total Organic Carbon). Based on Table 2, a concentration of 1.000 mg/µL was determined from the SSM-TOC analysis, indicating the amount of organic carbon in the fibreboard sludge sample. The computed SSM-TOC percentage of 23.39% indicates that the fibreboard sample contained a considerable amount of organic carbon. The TOC result of 23.39% was categorized as a high organic content in the sludge [14]. Depending on the source and method of treatment, sludge was classified as high range when it ranged within 15% to 25% [14]. TOC values in the 20% to 28% range, which were higher, indicated that the sludge was better suited for organic fertilizer applications or composting [15].

Table 2. Summaries of results for TOC analysis

Type	Analysis	Manual Dilution	Density	Result
Unknown	SSM-TOC	1.000	1.000 mg/µL	SSM-TOC: 23.39% SSM-TC: 23.40% SSM-IC: 0,00959%

Table 3. Summary of SSM-TC analysis

Area	CNV	Abs C	Concentration	SD Concentration	Weight	Volume
1400	1400	7020 µg	23.40%	0.000	30.00 mg	30 µL

Table 4. Summary of SSM-IC analysis

Area	CNV	Abs C	Concentration	SD Concentration	Weight	Volume
0.4755	0.4755	2.897 µg	0.00959%	0.000	30.20 mg	30 µL

However, the total carbon content of the fibreboard sample, as determined by the SSM-TC analysis, was 23.40% (Table 3). The consistency of the carbon content across measurements is suggested by this finding and the lack of a large standard variation in the concentration.

On the other hand, Table 4 shows that the SSM-IC analysis revealed the presence of inorganic carbon molecules at a low concentration of 0.00959%. The small standard deviation adds further evidence of the validity of this finding.

https://doi.org/10.31436/iiumej.v27i2.3858

From the TOC results, the higher range of the organic carbon content in the fibreboard sludge needed to be remediated to ensure it could reduce the potential environmental risks and be safe to be disposed. The presence of organic content in Portland cement is generally negligible, as the material is primarily composed of inorganic constituents, including iron oxide, limestone, silica, and alumina [16]. Any organic contaminants were burnt out during the high-temperature kiln process used in Portland cement production [16]. Furthermore, research indicated that the hydration and strength development of cemented soil can be inhibited by the presence of organic matter [17]. Other studies have also shown that an increase in organic matter weakens the mechanical properties of treated soils by interfering with cementitious binding, resulting in reduced strength and greater susceptibility to cracking [18].

3.2. Concentration of Heavy Metals in Raw Materials

The concentrations of heavy metals in the raw materials, including Portland cement and fibreboard sludge, are shown in Table 5. The concentration of heavy metals in each raw material was determined by averaging triplicate samples. The information presented allows for comparison of the concentrations of heavy metals (Cr, Cu, Zn, As, and Pb) in Portland cement and fibreboard sludge with USEPA standards. Waste exceeding the regulatory limit is classified as hazardous and must be properly managed.

Table 5. Concentration of heavy metals in raw materials

Heavy Metals	Formula	Concentration (mg/L)		Standard USEPA limit (mg/L) [19]
		Portland Cement	Fibreboard Sludge	
Chromium	Cr	36	4.06	≤ 0.1
Copper	Cu	8	3.79	≤ 1.3
Zinc	Zn	81	87.60	≤ 5.0
Arsenic	As	19	0.79	≤ 0.01
Lead	Pb	15	2.52	≤ 0.02

The results showed that all heavy metal elements in the Portland cement and fibreboard sludge exceeded the USEPA standard. Zn showed the highest concentration in the Portland cement and fibreboard sludge, with 81 mg/L and 87.60 mg/L, respectively. Therefore, remediation was necessary to reduce the toxicity of these harmful metal components in the environment. Mixing Portland cement as a binder with fibreboard sludge in the S/S method treatment may reduce heavy metal leaching. S/S method is an immobilization technique that uses binders to change the mobile metal (loid)s in the sludge into less mobile forms [20]. The study discovered that in polluted soil, Portland cement, fly ash, quicklime, and blast furnace slag can be used as composite materials to significantly stabilize heavy metals such as Cu, Cr, Zn, and Pb [21]. Fixation of heavy metals can be improved by adding more cement [3].

This research implemented the combination of S/S and bioremediation (bioaugmentation) to immobilize both heavy metals and organic contaminants. The combination treatment method was used because contaminant substances remain in the matrices and are not destroyed, posing a serious threat to the safety of solidified samples when the binder does not absorb or react with them to form a stable component. The discussion on the Chemical Oxygen Demand (COD) concentration for the combination method is discussed in the next section.

https://doi.org/10.31436/iiumej.v27i2.3858

3.3. Comparison of Chemical Oxygen Demand (COD) Concentration for Immediate Inoculation and Pre-Inoculation

Due to the high organic content of the sludge, the COD concentration was selected to indirectly quantify organic contaminants in the leachate. The test will show the mass of oxygen consumed per Liter of solution. COD measures the amount of oxygen required to oxidize contaminants. The efficiency of the combination method (S/S) and bioaugmentation in reducing COD concentration was demonstrated in Figure 7. Based on data from 14 days, immediate-inoculation and pre-inoculation for both fungi and bacteria yielded better results than the control sample. In the control sample, only indigenous microorganisms are reducing the COD concentration in the leachate. Based on this result, the indigenous microorganisms are unable to reduce the COD concentration. The potential of fungal and bacterial inoculation can help reduce COD concentration.

By comparing fungi and bacteria, fungi's immediate inoculation shows the lowest COD concentration, which means fungi can work immediately during the mixing process with cement. In contrast, bacteria require an adaptation period to utilize COD for their vital physiological and biochemical processes. Unbalanced nutrient levels and unfavorable conditions, such as high pollutant concentrations, temperature, moisture content, and pH, may affect the success of bacterial pollutant degradation [22]. Fungi are filamentous organisms that can break down complex organic pollutants, such as phenolic compounds, as soon as they encounter them and can react quickly to the presence of contaminants [23,24]. Fungi have demonstrated their immediate usefulness in bioremediation applications by degrading pharmaceutically active chemicals (PhACs) through complex enzymatic systems without major lag [25].

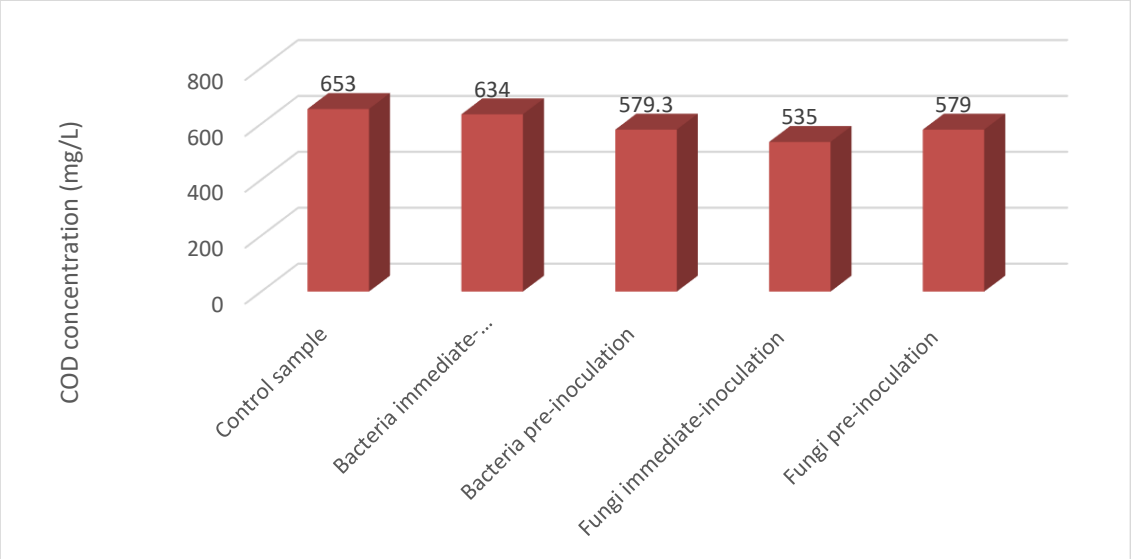


Figure 7. COD concentration (mg/L) for immediate-inoculation and pre-inoculation on 14 days of the curing period.

The experiment was extended to a 28-day curing period, and the focus was on immediate inoculation with fungi and bacteria, showing an increase in COD concentrations from 14 to 28 days. The control sample also shows the same trend, with COD concentrations increasing from 14 days to 28 days. Only samples with pre-inoculation show a reduction of COD concentrations over a longer duration. Microorganisms that fail to survive in S/S matrices under non-indigenous conditions may necessitate immediate inoculation. The organic content of the mixture increases due to the presence of dead microorganisms. This statement was supported

https://doi.org/10.31436/iiumej.v27i2.3858

by [26], which found that the remains of dead microorganisms contribute to soil organic matter, indicating that these residues are essential for increasing soil organic carbon content.

However, the results of the pre-inoculation method showed effectiveness: after 28 days, pre-inoculation with bacteria and fungi resulted in lower COD concentrations than those observed with immediate inoculation. It was shown that the adaptability of these bio-agents under the early 7-day inoculation condition (pre-inoculation) enhanced their ability to degrade organic contaminants, thereby reducing the COD concentration, as shown in Figure 8.

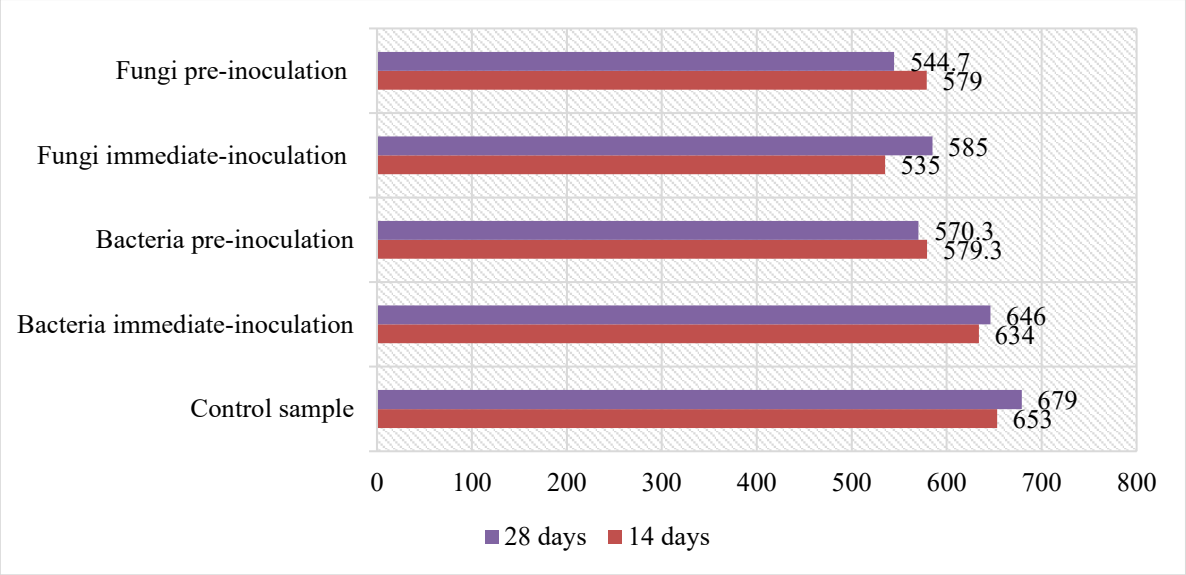


Figure 8. Comparison of COD concentration (mg/L) for immediate-inoculation and pre-inoculation between 14-day and 28-day curing periods.

In addition, pre-inoculation with fungi shows better results than with bacteria. It was implied that the ability of *Sphingobacterium spiritivorum* to break down the organic contaminants in the fibreboard sludge might be limited. These conditions demonstrated that the adaptability of the indigenous microorganism was lower than that of Sphingobacterium spiritivorum and Aspergillus brasiliensis during pre-inoculation.

The selection of the bio-agents greatly impacted the rates of the COD concentration reduction. The research conducted by [27] showed that the microbiological species involved were essential for the replication and enhancement of the bioaugmentation process. The microbial population's variety and functionality contributed to the repair of contaminated environments [28, 29].

In conclusion, comparing the reduction in the COD concentration between the immediate-inoculation and pre-inoculation procedures emphasized the importance of introducing the bio-agents in a timely manner when using the combination method of S/S and bioaugmentation. Based on this study, pre-inoculation appeared as the recommended strategy for lowering COD concentration.

3.4. Comparison of Heavy Metals Concentration for Immediate Inoculation and Pre-Inoculation

Because heavy metals in Portland cement and fibreboard sludge exceeded the USEPA standard, this parameter was also measured in the leaching test. In this study, the concentrations of heavy metals, including chromium (Cr), copper (Cu), zinc (Zn), arsenic (As), and lead (Pb) were measured and compared using the combination method, namely S/S and bioaugmentation

https://doi.org/10.31436/iiumej.v27i2.3858

method, with two inoculation methods, which were immediate-inoculation and pre-inoculation after 14 days of curing period, as shown in Figure 9.

The results in Figure 9 demonstrated that the pre-inoculation sample showed a greater reduction in heavy metals than the immediate-inoculation samples. All heavy metal concentrations were increased for the immediate-inoculation sample compared to the control sample, except for Cr (bacteria immediate-inoculation). Based on this finding, the combination method is not recommended for immediate inoculation. The bioaugmentation process in the S/S matrices may be inhibited by the presence of cement. Microbial life can be adversely affected by cement, which can change the chemical and physical environment [30,31]. Furthermore, a crucial part of the bioaugmentation process, microbial growth and activity can be inhibited by high pH levels. The process of mixing sludge with cement increases the pH level in the S/S matrices.

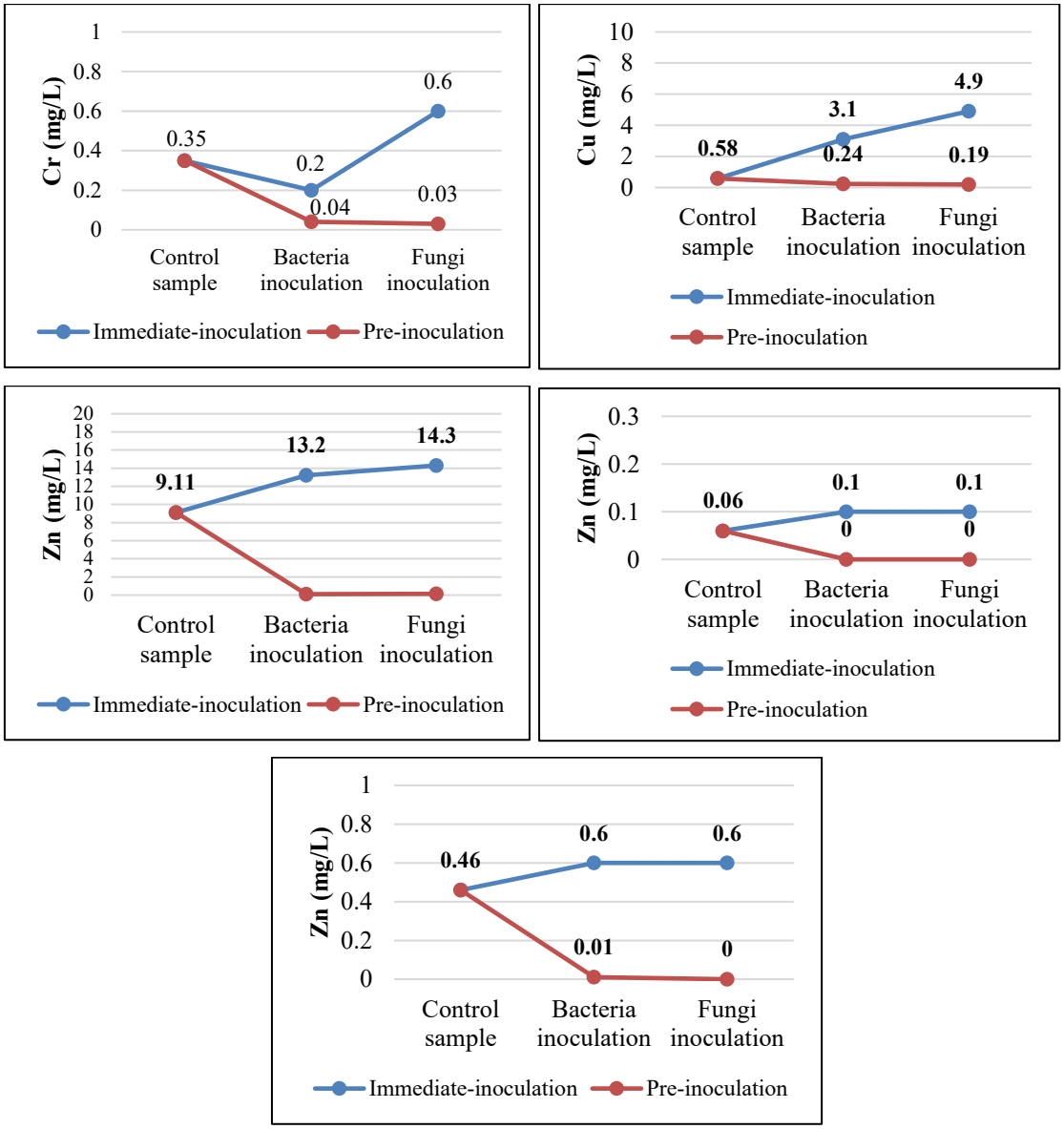


Figure 9. Heavy metals concentration (mg/L) for immediate-inoculation and pre-inoculation on 14 days of the curing period.

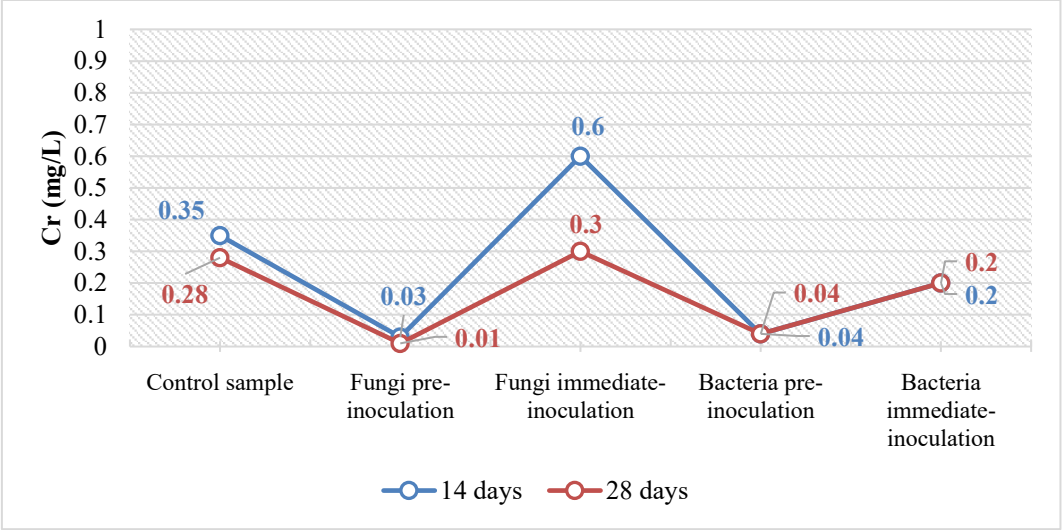
By comparison, pre-inoculation with fungi appeared to be the most effective method for reducing heavy metal concentrations, except for Zn. For Zn, pre-inoculation with bacteria

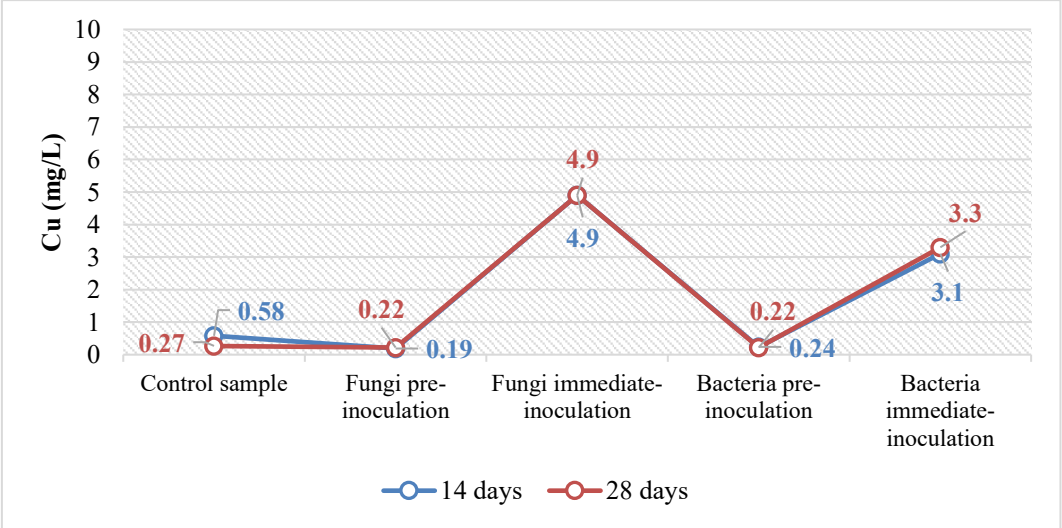
https://doi.org/10.31436/iiumej.v27i2.3858

shows a greater reduction in Zn concentration. Fungi are very useful as bioagents because they contain specialized enzymatic systems that can degrade various organic and inorganic contaminants [32]. The main components of these enzymatic systems are extracellular and intracellular enzymes, each with a specific function in the degradation process [23]. Laccase and peroxides were the key extracellular enzymes that catalyzed the breakdown of complex contaminants. Moreover, fungal bioremediation capacities are greatly enhanced by intracellular enzymes such as cytochromes [23].

Based on the results in Figure 10, the concentration of heavy metals decreased after a 28-day curing period, except for Cu and As, for both bacteria and fungi, following immediate inoculation. However, better results were obtained for pre-inoculation samples, in which As was not detected after 28 days of the curing period. Whereas for Cu, a slight reduction was observed with bacterial pre-inoculation and a slight increase with fungal pre-inoculation. For Cr, Zn, and Pb, the pre-inoculation showed a greater decrease in the concentration of heavy metals after 28 days compared to immediate-inoculation.

The concentration of heavy metals in fibreboard sludge treated with bacteria and fungi decreases clearly during this time, indicating that the early formation of microbial consortium speeds up the bioaugmentation process and leads to more noticeable decreases in the heavy metal's concentration over time.





https://doi.org/10.31436/iiumej.v27i2.3858

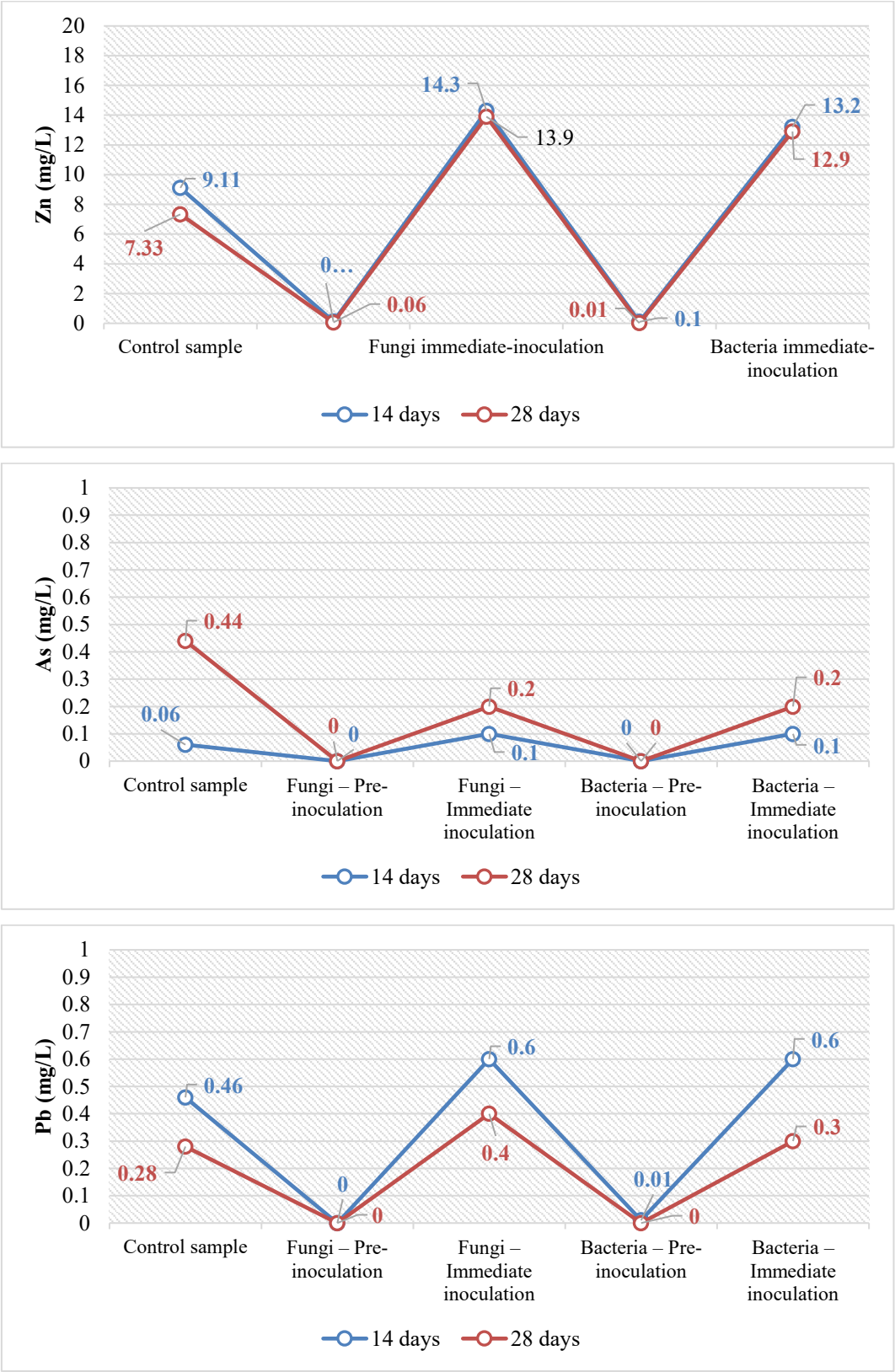


Figure 10. Comparison of heavy metals (mg/L) concentration between 14 and 28 days of curing

3.5. Statistical Comparison of Immediate Inoculation and Pre-Inoculation

The bar chart utilizes a logarithmic scale to effectively illustrate the substantial differences in values among the various parameters that have been developed. This approach allows for a

https://doi.org/10.31436/iiumej.v27i2.3858

clearer comparison, particularly when the values vary widely in magnitude. As depicted in Figure 11, the logarithmic scale enables the visualization of these differences in a more manageable way, making it easier to interpret the data.

Independent t-tests were conducted to analyze the effects of immediate inoculation relative to pre-inoculation across a range of parameters, as presented in Table 6. The results showed statistically significant differences in three key parameters: COD ($p = 0.013$), Cu ($p = 0.001$), and Zn ($p = 0.0001$). Moreover, no significant differences were observed for Cr ($p = 0.08$), As ($p = 0.095$), and Pb ($p = 0.097$), as the p-values are greater than 0.05.

These findings indicate that the immediate-inoculation treatment generally resulted in higher concentrations of all measured parameters than the pre-inoculation treatment. Notably, the most pronounced differences were observed for COD, Cu, and Zn, suggesting that pre-inoculation may have a substantial positive effect on these specific parameters. This suggests the potential benefits of pre-inoculation in reducing the concentrations of these critical substances.

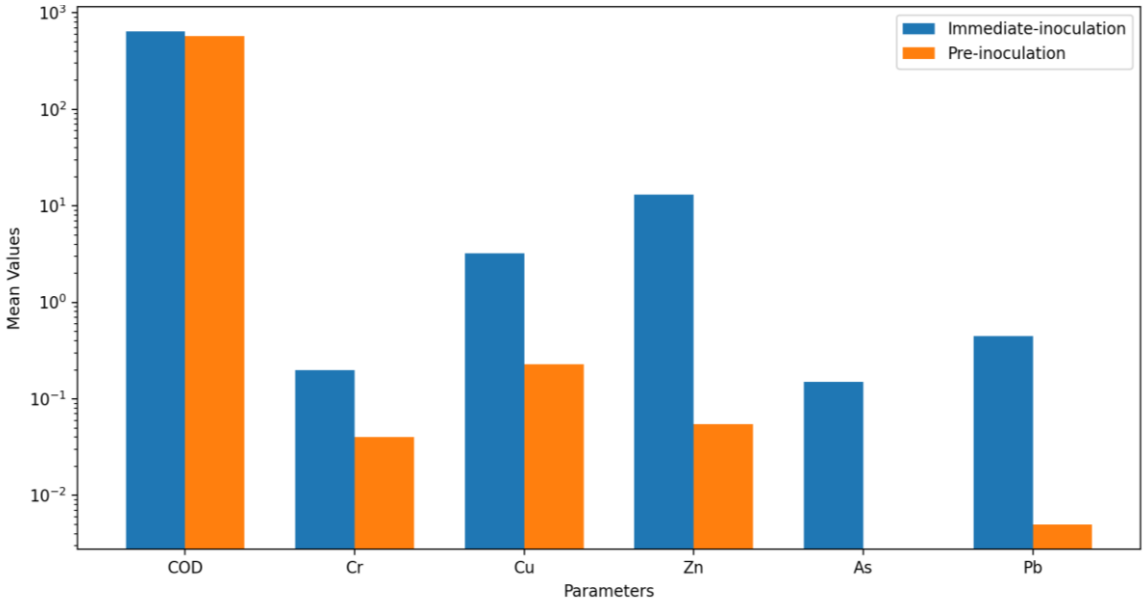


Figure 11. Comparison of parameter concentrations on a logarithmic scale

Table 6. Result of t-test analysis across all parameters

Parameter	Mean		t-statistic	p-value	Significance
	Immediate Inoculation	Pre-Inoculation			
COD	640.0	574.8	8.6933333333	0.0129750861	*
Cr	0.2	0.04	0.0001	0.08	***
Cu	3.2	0.23	29.5526045492	0.0011430451	*
Zn	13.05	0.055	82.979683843	0.0001451984	*
As	0.15	0.0	3.0	0.0954659663	***
Pb	0.45	0.005	2.9650198907	0.0974113588	***

**p ≤ 0.01: Highly Significant, **p ≤ 0.05: Significant, ***p > 0.05: Not Significant*

COD, Cu, and Zn, selected as key parameters, were thoroughly analyzed to evaluate and compare the roles and effects of bacteria versus fungi in the inoculation method. The Shapiro-Wilk test results indicate that the datasets for Zn and Cu do not follow a normal distribution, as p-values fall below the 0.05 threshold, as shown in Table 7. This finding suggests that

parametric tests, which assume normality, would not be appropriate for analyzing these data. As a result, the Mann-Whitney U test was applied as a nonparametric statistical method that does not rely on the assumption of normality to compare the presence of bacteria and fungi across three specific parameters.

Upon conducting the Mann-Whitney U test, the ranks of the data for both bacteria and fungi were analyzed. As shown in Table 8, p-values from this analysis indicate no statistically significant differences between the two groups (bacteria and fungi) in Zn and Cu measurements. This finding suggests that the presence of either bacteria or fungi does not significantly influence the levels of these parameters. Therefore, any observed variations in COD, Zn, and Cu levels between the two groups may be attributed to random variability rather than a systematic difference.

Table 7. Statistical comparison between bacteria and fungi using the Shapiro-Wilk test

	Shapiro-Wilk test		
	COD	**Zn**	**Cu**
Bacteria	0.2470*	0.0323**	0.0479**
Fungi	0.3046*	0.0331**	0.0255**

**p > 0.05: t-test acceptable, **p < 0.05: Use Mann-Whitney U*

Table 8. Mann-Whitney U Test results for Cu and Zn concentrations (bacterial treatment)

	Mann-Whitney U Test	
	Zn	**Cu**
Bacteria	0.4857**	1.0000**
Fungi		

**p ≤ 0.05: significant difference, **p > 0.05: not significant*

The differences between the two inoculation methods, namely immediate inoculation and pre-inoculation, underscore the critical importance of carefully selecting the optimal timing for inoculation before applying the binder in the stabilization/solidification (S/S) method. Immediate inoculation involves adding microbes directly into the sludge at the time of mixing, while pre-inoculation requires preparing the microbes in advance and allowing them to establish themselves before binder application. This timing is essential, as it can significantly affect the effectiveness of microbial interactions with the sludge. However, this study found no significant differences in effectiveness between the bacterial and fungal species used in these inoculation methods. This lack of significance suggests that both types of microorganisms can be used interchangeably without affecting the overall outcomes of the S/S process. Based on the findings of this study, the inoculation method for enhancing contaminant removal in cement matrices contributes new knowledge to the sludge remediation process.

4. DISCUSSION

The results of this study indicate that pre-inoculation with both fungi and bacteria markedly improves the stabilization of COD and heavy metals compared with immediate inoculation. This enhanced performance is clearly illustrated in the comparative analyses presented in Figures 12 and 13, which highlight the superior effectiveness of pre-inoculated treatment.

https://doi.org/10.31436/iiumej.v27i2.3858

As shown in Figure 12, pre-inoculation resulted in greater reductions of COD and heavy metals than immediate inoculation after 14 days, as reflected by the percentage contributions of each treatment. For COD, pre-inoculation accounted for approximately 55–60% of the total reduction, exceeding the contribution of immediate inoculation and indicating enhanced organic matter degradation following microbial acclimatization. A similar trend was observed for heavy metals, where pre-inoculation consistently contributed the largest percentage of removal across all metals studied. Bacteria pre-inoculation accounted for more than 60% of chromium (Cr) and copper (Cu) reduction, while fungal pre-inoculation contributed nearly 70% of zinc (Zn) removal. For arsenic (As) and lead (Pb), pre-inoculated systems contributed over 65% of the total reduction, whereas immediate inoculation contributed comparatively little. Overall, the dominance of pre-inoculation in percentage reduction for both COD and heavy metals indicates that microbial adaptation prior to exposure significantly improves treatment efficiency. These findings highlight pre-inoculation as a more effective strategy for achieving rapid contaminant reduction within short treatment durations.

At 28 days, the percentage contribution analysis, as depicted in Figure 13, shows that pre-inoculation resulted in greater reductions in COD and heavy metals than immediate inoculation. For COD, pre-inoculated treatments accounted for approximately 55–60% of the total reduction, indicating improved organic matter stabilization following microbial adaptation. For heavy metals, pre-inoculation consistently dominated the percentage contribution across all metals studied. Reductions in Cr and Cu were largely driven by bacterial and fungal pre-inoculation, accounting for more than 40–50% of total removal, whereas Zn reduction exceeded 60% under pre-inoculated conditions. Similarly, As and Pb removal was predominantly associated with pre-inoculation, accounting for over 65% of the total reduction, whereas immediate inoculation contributed comparatively little. These results confirm that microbial pre-conditioning enhances long-term contaminant stabilization, making pre-inoculation a more effective treatment strategy over extended durations.

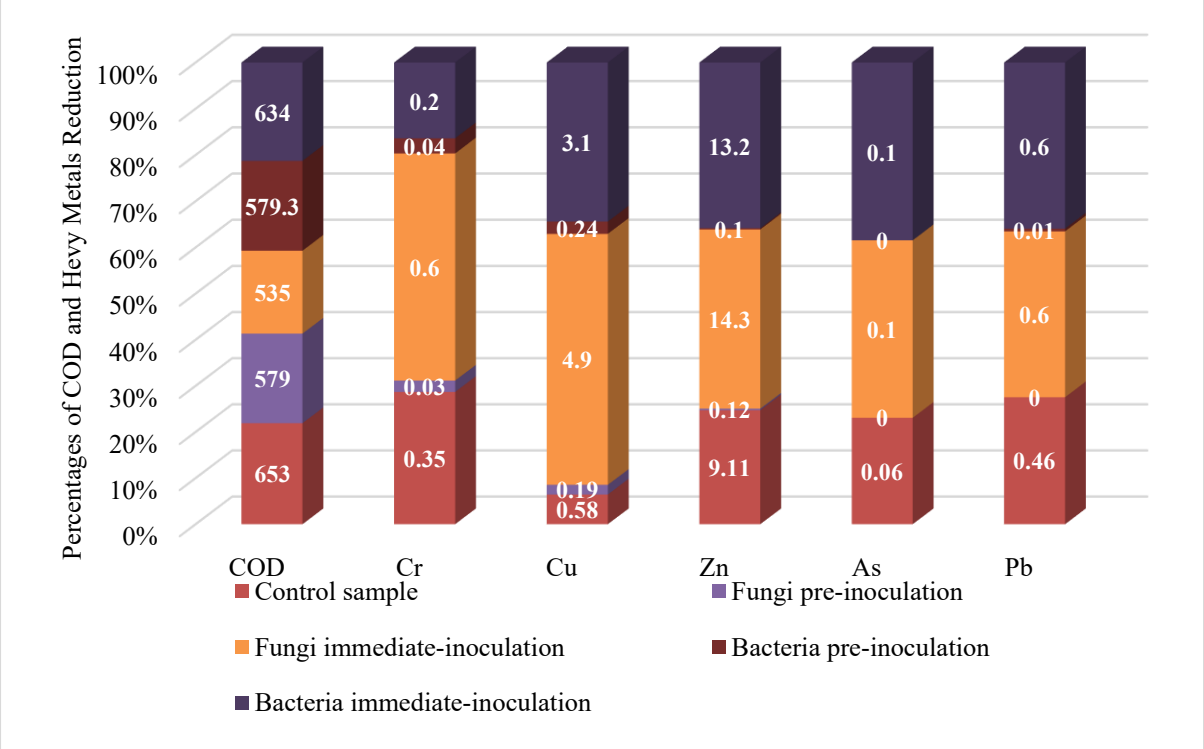


Figure 12. Comparison of the percentage reduction for COD and heavy metals in 14 days

https://doi.org/10.31436/iiumej.v27i2.3858

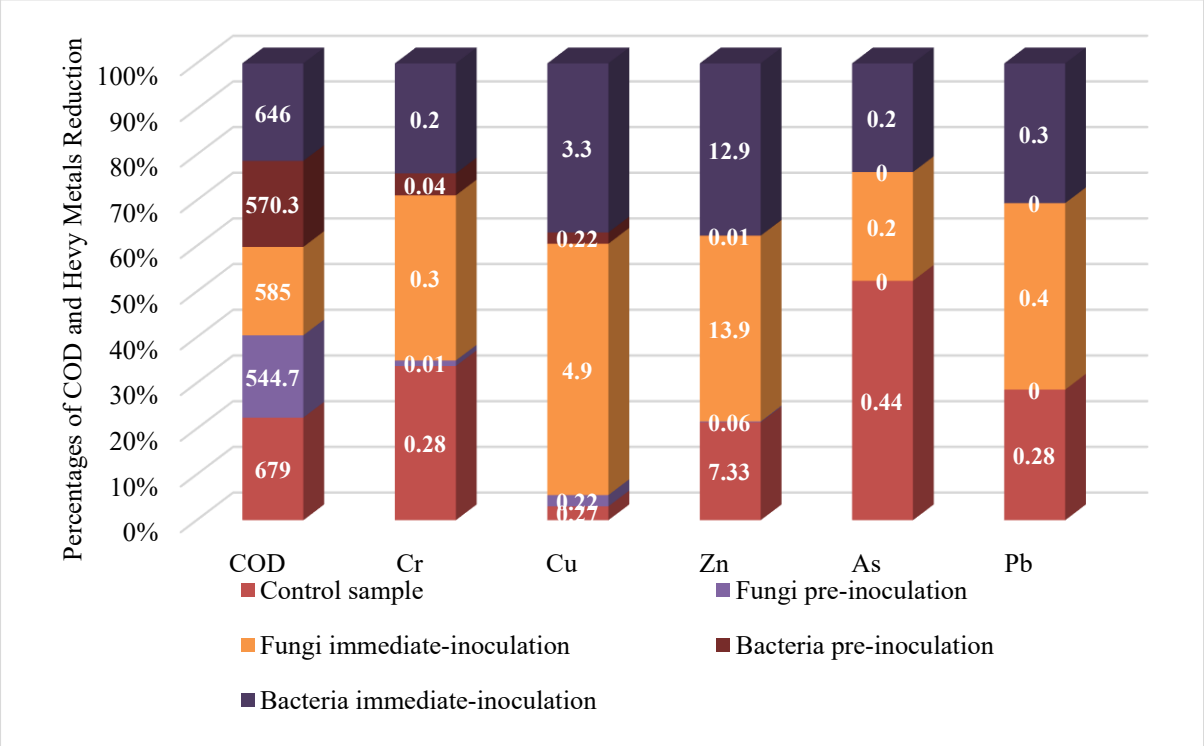


Figure 13. Comparison of Percentage Reduction for COD and Heavy Metals in 28 Days

The stabilization performance of pre-inoculation and immediate inoculation S/S treatments is presented in Table 9, where COD, Cu, and Zn were identified as the most influential contaminants based on statistical significance analysis. Focusing on these key parameters provides a clearer understanding of treatment effectiveness. Consistent with the percentage-based reduction analysis, pre-inoculation contributed a larger share of the reductions in COD and heavy metals than immediate inoculation, particularly for COD, Cu, and Zn. The higher percentage reductions observed in pre-inoculated systems confirm that microbial pre-conditioning enhances the efficiency of contaminant stabilization, supporting its effectiveness as a superior S/S treatment strategy.

Table 9. COD, Cu, and Zn levels for immediate inoculation and pre-inoculation (14 and 28 days)

Treatment	14 days			28 days		
	COD (mg/L)	Cu (mg/L)	Zn (mg/L)	COD (mg/L)	Cu (mg/L)	Zn (mg/L)
Fungi pre-inoculation	579	0.19	0.12	544.7	0.22	0.06
Fungi's immediate inoculation	535	4.90	14.30	585	4.90	13.90
Bacteria pre-inoculation	579.3	0.24	0.10	570.3	0.22	0.01
Bacteria's immediate inoculation	634	3.10	13.20	646	3.30	12.90
Control samples	653	0.58	9.11	679	0.27	7.33

COD reduction at 28 days was more pronounced in pre-inoculated samples (fungi: 544.7 mg/L; bacteria: 570.3 mg/L) than in immediate-inoculated samples (fungi: 579.0 mg/L; bacteria: 579.3 mg/L), suggesting that allowing microbial populations to establish prior to treatment accelerates organic matter degradation. Heavy metal immobilization, particularly for Cu and Zn, was markedly improved under pre-inoculation at 28 days, with Cu concentrations

as low as 0.22 mg/L for fungi and bacteria, Zn concentrations were reduced to 0.06 mg/L in fungi and 0.01 mg/L in bacteria under pre-inoculation, whereas under immediate inoculation, Cu and Zn levels remained substantially higher (Cu: fungi 4.90 mg/L, bacteria 3.30 mg/L; Zn: fungi 13.90 mg/L, bacteria 12.90 mg/L). Control samples consistently exhibited the highest COD and intermediate metal levels, highlighting the contribution of microbial inoculation to both organic matter breakdown and metal stabilization. These results highlight the practical advantage of pre-inoculation for sludge remediation, as both organic and inorganic contaminants can be effectively stabilized prior to disposal or reuse.

Pre-inoculation treatment in the S/S method developed in this study demonstrates significant improvements over prior fungal bioaugmentation studies. As shown in Table 10, previous studies have investigated fungal strategies for soil remediation and heavy metal removal, and filamentous fungal consortia applied to landfill and contaminated soils exploit synergistic interactions to enhance pollutant degradation and metal immobilization [34, 35]. Indigenous soil fungi offer a sustainable remediation approach with minimal ecological impact [36]. This non-living fungal biomass and liquid-phase treatments have also been used to remove heavy metals via extracellular binding.

These studies predominantly relied on simultaneous inoculation or non-S/S matrices, which can delay microbial adaptation and reduce stabilization efficiency. In contrast, the pre-inoculation approach in this study allowed microbial communities, including *Aspergillus brasiliensis* and *Sphingobacterium spiritivorum*, to establish metabolic activity before exposure to the sludge matrices. This led to accelerated COD reduction and markedly lower residual concentrations of Cu and Zn. To our knowledge, this is the first study to integrate pre-inoculated microbes into S/S matrices for simultaneous organic and heavy-metal stabilization, emphasizing both the novelty and industrial scalability of the approach, as shown in Table 10.

Table 10. Comparison of previous fungal bioaugmentation studies and the present pre-inoculation S/S approach.

Study	Fungi	Method Approach	Key Findings	Novelty
Hassan *et al.*, 2020 [34]	Filamentous fungal consortia	Simultaneous inoculation	Metal removal (Cu ~52%) over long-term incubation	Did not integrate fungi into S/S; slower adaptation
Hassan *et al.*, 2020 [35]	Fungal consortia	Simultaneous inoculation	Enhanced metal removal vs. control	Soil-based; no pre-inoculation; not S/S
Dusengemungu *et al.,* 2022 [36]	Indigenous fungi (*Aspergillus, Cladosporium, Geotrichum*)	Soil bioaugmentation	High Zn removal; metal tolerance	Soil system; not cemented; no pre-inoculation
Current study (2025)	*Aspergillus brasiliensis* + *Sphingobacterium spiritivorum*	Pre-inoculation	High reductions of COD and heavy metals	First to integrate pre-inoculated microbes into S/S matrices

From a practical perspective, these findings have clear implications for industrial waste treatment. Pre-inoculation could be integrated into S/S processes at water treatment plants, industrial sludge management facilities, or contaminated site remediation projects. By establishing active microbial populations beforehand, treatment efficiency can be improved, potentially reducing curing times and lowering the residual toxicity of treated sludge.

Furthermore, the methodology can be scaled to larger systems by pre-culturing microbial inocula in controlled bioreactors before mixing with bulk sludge, offering a feasible route for industrial-scale application.

Overall, this study not only demonstrates the technical benefits of pre-inoculation over immediate inoculation but also provides a framework for applying microbial S/S strategies in real-world sludge remediation scenarios. Future work should explore optimization of inoculum size, curing conditions, and microbial consortia combinations to maximize treatment efficiency across diverse sludge types and heavy metal compositions.

5. CONCLUSIONS

According to the TOC analysis, the sludge contained 23.39% organic carbon, indicating a high organic content in the fibreboard sludge. The heavy metals present in both the sludge and Portland cement included Cr, Cu, Zn, As, and Pb, all of which exceeded USEPA standards. After 28 days, the control sample and the pre-inoculated bacteria and fungi showed lower COD concentrations than the immediate-inoculated samples. When comparing the COD concentrations of bacteria and fungi, the fungi with pre-inoculation exhibited the lowest COD concentration at 544.7 mg/L after 28 days, while bacteria with pre-inoculation had a concentration of 570.3 mg/L after the same period. Additionally, based on the heavy metal analysis, the pre-inoculation samples showed a greater reduction in heavy metal concentrations than the immediate-inoculation samples. Specifically, the pre-inoculation with fungi had the most substantial impact on lowering heavy metal concentrations, achieving levels of 0.01 mg/L for Cr, 0.22 mg/L for Cu, 0.06 mg/L for Zn, and 0 mg/L for both As and Pb after 28 days. Statistically proven that pre-inoculation is better than immediate-inoculation with a significant p-value of 0.0001 for COD, Cu, and Zn. The results from the pre-inoculation method were particularly noteworthy. However, there is no significant difference between the two microbes used in this study. Pre-inoculation with a co-culture of bacteria and fungi is recommended to further enhance the solid-state treatment method, providing a promising approach for industrial sludge management. Future research is required to evaluate scalability for practical and sustainable waste treatment applications.

ACKNOWLEDGEMENT

This study was supported by Universiti Tun Hussein Onn Malaysia (UTHM) through GPPS & Publisher's Office (vot Q776 and E15216).

REFERENCES

[1] Ibrahim, M. F., Hod, R., Toha, H. R., Mohammed Nawi, A., Idris, I. B., Mohd Yusoff, H., & Sahani, M. (2021). The impacts of illegal toxic waste dumping on children's Health: A Review and Case Study from Pasir Gudang, Malaysia. International Journal of Environmental Research and Public Health, 18(5), 2221.

[2] Oluseun Adejumo, I., & Adebukola Adebiyi, O. (2021). Agricultural Solid Wastes: Causes, effects, and effective management. Strategies of Sustainable Solid Waste Management.

[3] Kogbara, R. B., Ayotamuno, J. M., Onuomah, I., Ehio, V., & Damka, T. G. D. (2016). Stabilisation/Solidification and bioaugmentation treatment of petroleum drill cuttings. Applied Geochemistry, 71, 1–8.

[4] Ding, X., Wang, J., Huang, Q., Hu, S., Wu, Y., & Wang, L. (2021). The effects of waste cement on the bioavailability, mobility, and leaching of cadmium in soil. International Journal of Environmental Research and Public Health, 18(16), 8885.

[5] Leonard, S. A., & Stegemann, J. A. (2010). Stabilization/Solidification of petroleum drill cuttings. Journal of Hazardous Materials, 174(1-3), 463–472.

[6] Jaishankar, M., Tseten, T., Anbalagan, N., Mathew, B. B., & Beeregowda, K. N. (2014). Toxicity, mechanism and health effects of some heavy metals. Interdisciplinary Toxicology, 7(2), 60–72.

[7] Boutammine, H., Salem, Z., & Khodja, M. (2020). Petroleum drill cuttings treatment using stabilization/solidification and biological process combination. Soil and Sediment Contamination: An International Journal, 29(4), 369–383.

[8] Ball, A. S., Stewart, R. J., & Schliephake, K. (2011). A review of the current options for the treatment and safe disposal of drill cuttings. Waste Management & Research: The Journal for a Sustainable Circular Economy, 30(5), 457–473.

[9] Cerqueira, V. S., Peralba, M. do, Camargo, F. A. O., & Bento, F. M. (2014). Comparison of bioremediation strategies for soil impacted with petrochemical oily sludge. International Biodeterioration & Biodegradation, 95, 338–345.

[10] Mohd-Kamil, N. A. F., Hussain, N. H., Mizad, M. B., & Abdul-Talib, S. (2014). Enhancing performance of Sphingobacterium spiritivorum in bioremediation phenanthrene contaminated sand. Remediation Journal, 24(3), 119-128.

[11] Petkova, K. (2013). Potencial of Aspergillus niger in bioremediation of contaminated soils. 13th SGEM GeoConference on ECOLOGY, ECONOMICS, EDUCATION AND LEGISLATION.

[12] Sharma, S., & Malaviya, P. (2016). Bioremediation of tannery wastewater by aspergillus flavus SPFT2. International Journal of Current Microbiology and Applied Sciences, 5(3), 137–143.

[13] Mohd-Rizal, S., Nasuha Mohd Rais, N., Filzah Mohd Kamil, N. A., Hamzah, N., Azliza Akbar, N., & Shaylinda Mohd Zin, N. (2022). Remediation of organic contaminant in sludge by using bioaugmentation and solidification and stabilization (S/S) method. IOP Conference Series: Earth and Environmental Science, 1022(1), 012061.

[14] Gaur, R. Z., Khoury, O., Zohar, M., Poverenov, E., Darzi, R., Laor, Y., & Posmanik, R. (2020). Hydrothermal carbonization of sewage sludge coupled with anaerobic digestion: Integrated approach for Sludge Management and Energy Recycling. Energy Conversion and Management, 224, 113353.

[15] Myszura-Dymek, M. (2024). Properties of organic matter in composts based on Sewage Sludge. Journal of Ecological Engineering, 25(7), 70–81.

[16] Lavagna, L., & Nisticò, R. (2022). An insight into the chemistry of cement—a review. Applied Sciences, 13(1), 203.

[17] Shao, L., Ding, Z., Wang, S., Pan, K., & Hu, C. (2023). Effect of organic matter components on the mechanical properties of cemented soil. Materials, 16(17), 5889.

[18] Zhong, Y., Cai, G., Wang, S., Qin, H., Zhang, C., & Li, J. (2022). Influence of organic content on the mechanical properties of organic-rich soils stabilized with Cao-ggbs binder and PC. Water, 14(19), 3053.

[19] Patra, S., Whaung, S. T., & Kwan, W. L. (2017). Analysis of heavy metals in incineration bottom ash in Singapore and potential impact of pre-sorting on ash quality. Energy Procedia, 143, 454–459.

[20] Oh, S., & Shin, W. S. (2017). Applicability of solidified/stabilized dye sludge char as a landfill cover material. KSCE Journal of Civil Engineering, 21(7), 2573–2583.

[21] Li, S., Yang, J., & Zhang, P. (2022). Hydration and hardening properties of reactive magnesia and Portland cement composite. Construction and Building Materials, 327, 126779.

[22] Smith, E., Thavamani, P., Ramadass, K., Naidu, R., Srivastava, P., & Megharaj, M. (2015). Remediation trials for hydrocarbon-contaminated soils in arid environments: Evaluation of bioslurry and biopiling techniques. International Biodeterioration & Biodegradation, 101, 56–65.

[23] Amobonye, A., Aruwa, C. E., Aransiola, S., Omame, J., Alabi, T. D., & Lalung, J. (2023). The potential of fungi in the bioremediation of pharmaceutically active compounds: A comprehensive review. Frontiers in Microbiology, 14.

[24] Torres-Farradá, G., Thijs, S., Rineau, F., Guerra, G., & Vangronsveld, J. (2024). White rot fungi as tools for the bioremediation of xenobiotics: A Review. Journal of Fungi 10(3), 167.

[25] Olicón-Hernández, D. R., Camacho-Morales, R. L., Pozo, C., González-López, J., & Aranda, E. (2019). Evaluation of diclofenac biodegradation by the ascomycete fungus Penicillium oxalicum at flask and bench bioreactor scales. Science of The Total Environment, 662, 607–614.

[26] Liang, C., Amelung, W., Lehmann, J., & Kästner, M. (2019). Quantitative assessment of microbial necromass contribution to soil organic matter. Global Change Biology, 25(11), 3578–3590.

[27] Johnson, O. A., & Affam, A. C. (2018). Petroleum sludge treatment and disposal: A Review. Environmental Engineering Research, 24(2), 191–201.

[28] Bastida, F., Jehmlich, N., Lima, K., Morris, B. E. L., Richnow, H. H., Hernández, T., von Bergen, M., & García, C. (2016). The ecological and physiological responses of the microbial community from a semiarid soil to hydrocarbon contamination and its bioremediation using Compost Amendment. Journal of Proteomics, 135, 162–169.

[29] Dangi, A. K., Sharma, B., Hill, R. T., & Shukla, P. (2018). Bioremediation through microbes: Systems Biology and Metabolic Engineering Approach. Critical Reviews in Biotechnology, 39(1), 79–98.

[30] Li, Y., Zhao, J., & Zhang, Z. (2021). Implementing metatranscriptomics to unveil the mechanism of bioaugmentation adopted in a continuous anaerobic process treating cow manure. Bioresource Technology, 330, 124962.

[31] Tribedi, P., Goswami, M., Chakraborty, P., Mukherjee, K., Mitra, G., Bhattacharyya, P., & Dey, S. (2018). Bioaugmentation and biostimulation: A potential strategy for environmental remediation. Journal of Microbiology & Experimentation, 6(5).

[32] Vaksmaa, A., Guerrero-Cruz, S., Ghosh, P., Zeghal, E., Hernando-Morales, V., & Niemann, H. (2023). Role of fungi in bioremediation of emerging pollutants. Frontiers in Marine Science, 10.

[33] Rais, N. N. M., Rachman, R. M., & Kamil, N. A. F. M. (2023). Potential of the Bioaugmentation Method to Enhance the Remediation Process in the Solidification/Stabilization Method. International Journal of Sustainable Construction Engineering and Technology, 14(5), 336-348.

[34] Hassan, A., Pariatamby, A., Ossai, I. C., & Hamid, F. S. (2020). Bioaugmentation assisted mycoremediation of heavy metal and/metalloid landfill contaminated soil using consortia of filamentous fungi. Biochemical Engineering Journal, 157, 107550.

[35] Hassan, A., Periathamby, A., Ahmed, A., Innocent, O., & Hamid, F. S. (2020). Effective bioremediation of heavy metal–contaminated landfill soil through bioaugmentation using consortia of fungi. Journal of Soils and Sediments, 20(1), 66-80.

[36] Dusengemungu, L., Gwanama, C., Simuchimba, G., & Mubemba, B. (2022). Potential of bioaugmentation of heavy metal contaminated soils in the Zambian Copperbelt using autochthonous filamentous fungi. Frontiers in Microbiology, 13, 1045671.

https://doi.org/10.31436/iiumej.v27i2.3387

Magnetic Resonance Imaging and Ultrasound-based Breast Tumor Classification using Soft Voting Ensemble Deep Learning Method

SURESH RAMAYANAM*, VIJAYA BATHINI, SANKAR BABU JANGILI, SANTOSH KUMAR REDDY THIKKAM

Department of Computer Science and Engineering, Sri Venkateswara College of Engineering, Tirupati, India

**Corresponding author: ramsuri42@gmail.com*

(Received: 6 August 2024; Accepted: 4 February 2026; Published online: 10 May 2026)

ABSTRACT: In recent years, breast cancer (BC) has been considered the most common cause of death in women worldwide. Commonly, BC occurs in breast cells or fatty tissues within the breast, and these tissues tend to grow faster and lead to death. The survival of patients with BC can be enhanced with early and accurate diagnosis. Therefore, this research proposes an ensemble learning approach that uses Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and a soft-voting mechanism to make early predictions and diagnoses of BC. The input sample images are initially passed through the convolutional layer to extract features from each modality separately. The Multi-Modality Adaptive Feature Fusion (MMAFF) technique combines generic features across modalities. Compared with existing techniques such as Deep Convolutional Neural Networks (DCNN) and Resolution Adaptive Network (RANet), the proposed ensemble learning technique achieves 89.6% accuracy for clinical indicator prediction and 99.25% accuracy for final tumor classification on the MRI-US dataset.

ABSTRAK: Dalam beberapa tahun kebelakangan ini, kanser payudara (Breast Cancer, BC) telah dikenal pasti sebagai antara punca utama kematian dalam kalangan wanita di seluruh dunia. Secara umum, BC berlaku pada sel-sel payudara atau tisu lemak dalam payudara, di mana sel-sel ini cenderung membiak dengan cepat dan berpotensi membawa kepada kematian. Kadar kelangsungan hidup pesakit BC dapat ditingkatkan melalui diagnosis awal yang tepat dan boleh dipercayai. Sehubungan itu, kajian ini mencadangkan satu pendekatan pembelajaran ansambel yang menggabungkan Rangkaian Neural Berulang (Recurrent Neural Networks, RNN), Rangkaian Neural Konvolusi (Convolutional Neural Networks, CNN), Memori Jangka Pendek Panjang (Long Short-Term Memory, LSTM), serta mekanisme pengundian lembut (soft voting) bagi tujuan ramalan awal dan diagnosis BC. Imej sampel input pada peringkat awal diproses melalui lapisan konvolusi untuk mengekstrak ciri daripada setiap modaliti secara berasingan. Seterusnya, teknik Multi-Modality Adaptive Feature Fusion (MMAFF) digunakan untuk menggabungkan ciri-ciri umum daripada setiap modaliti tersebut. Berbanding dengan teknik sedia ada seperti Rangkaian Neural Konvolusi Mendalam (Deep Convolutional Neural Networks, DCNN) dan Resolution Adaptive Network (RANet), pendekatan pembelajaran ansambel yang dicadangkan menunjukkan prestasi yang lebih baik dengan mencapai ketepatan sebanyak 89.6% bagi ramalan penunjuk klinikal dan 99.25% bagi pengelasan akhir tumor menggunakan set data MRI-US.

KEYWORDS: *Breast Tumor Classification, Dynamic Contrast-Enhanced Magnetic Resonance Imaging, Ensemble Deep Learning, Multi-Modality Adaptive Feature Fusion Framework, Ultrasound.*

1. INTRODUCTION

Breast Cancer (BC) remains a severe health concern among females since it accounts for a high percentage of cancer-related mortality worldwide [1]. So far, several studies have investigated the use of diverse medical imaging modalities to identify BC; however, the precision of each model varies with the available dataset and the nature of the problem [2]. There is significant dependence on breast cancer imaging, making the quality of the developed images paramount for enabling accurate pathological results [3]. Additionally, several other variables significantly impact the accuracy of pathology tests [4]. These include the clinicians' degree of experience and the amount of attention they dedicate to handling samples, among other factors [5]. The diagnosis of BC using deep learning models has been extensively researched because it can improve patient survival, reduce the disease burden, and advance healthcare technology [6]. Hence, there is a need to incorporate the learning model, given advances in computer and artificial intelligence technologies, to help doctors provide precise, faster breast cancer treatment [7-8].

Deep learning models have gained further adoption within the computer vision community, particularly in biomedical image analysis, where they provide a rich set of complementary features and automatically extract both high- and low-level features from images [9-10]. Convolutional neural networks, a type of deep learning, are applied to handle histories of breast cancer images, where the primary classification is into two classes [11-12]. CNNs have also been used to learn deep representations from BC images, which are then used in models built with other classical DL methods to classify features derived from images [13]. Performance analysis of CNNs compared with traditional classifiers for pathology image classification of elite breast cancer images suggests that the former has emerged as the best among them [14]. Breast pathology image analysis using semi- or fully automated methods helps pathologists achieve better results, optimize the diagnostic workflow, and speed it up. DL models capitalize on the small features in images that are hard to overlook, given their nature [15]. Such an analysis is motivated by three key challenges widely present in current breast tumor classification methods: (1) most breast tumor classification methods rely on single-modality imaging, which may compromise diagnostic reliability; (2) many multi-modality fusion techniques cannot effectively disentangle modality-specific from modality-agnostic features; and (3) individual deep learning models often exhibit unstable performance across heterogeneous clinical data. Therefore, in this research, the aim is to: (i) design a discrimination–adaptation mechanism for separating modality-specific and modality-agnostic features with adversarial learning; (ii) propose an MMAFF framework to improve the compactness and robustness of the learned features; and (iii) build a reliable breast tumor classifier using MRI-US information based on a soft-voting ensemble strategy incorporated with CNN, RNN, and LSTM networks.

The major contributions are listed as follows:

- The ensemble deep learning method is used for classification from MRI-US. The features are extracted from various modalities distinctly by 2D and 3D convolutional layers.

- A discrimination-adaptation part is presented to attain agnostic-specific features using adversarial learning with 3 distinct discriminators.

- The feature fusion module is proposed by using MAFF to combine the modality-agnostic features and increase feature solidity by exploiting an Affinity Matrix (AM) with Nearest Neighbor Selection (NNS).

The structure of this research is as follows: Section 2 analyzes the related works. Section 3 expands the proposed ensemble learning method, and Section 4 presents the results analysis and comparison. The conclusion of this research is demonstrated in Section 5.

2. LITERATURE REVIEW

The goal of breast tumor classification is to extract and identify tumors. The relevant research focused on DL techniques for segmenting and classifying liver tumors is presented here.

Kranti Kumar Dewangan et al. [16] presented a Back Propagation Boosting Recurrent Wienmed (BPBRW) with Hybrid Krill Herd African Buffalo Optimization (HKH-ABO) model for identifying BC. This was achieved by the optimal performance of the remaining models in terms of low energy consumption and precise severity analysis. However, the combination of multiple optimization techniques and models increases system complexity, making implementation harder and requiring a large number of parameters.

Muhammad Junaid Umer et al. [17] implemented a 6B-Net Deep CNN with a feature fusion and selection model for the multi-class classification of BC. This model was efficiently applied to early detection, using Particle Swarm Optimization (PSO) and Ant Colony Search (ACS) for feature selection before performing classification. These methods converge faster and often require fewer iterations to achieve accurate breast tumor classification, effectively finding optimal paths and solutions. However, these optimization techniques struggle with high-dimensional optimization, leading to large-scale issues due to the need to update pheromone trails and evaluate multiple solutions.

Muhammad Junaid Umer et al. [18] proposed a multi-class breast cancer classification framework that incorporates 6B-Net with deep feature fusion and a selection approach. Thereby, the recommended framework comprises three phases: firstly, a 35-layer deep CNN model with a simultaneous processing block was demonstrated. Subsequently, to enable effective feature learning, the recommended approach was initially trained on the CIFAR-100 dataset. Therefore, the model was used as a feature extractor and then efficiently pre-trained for multi-class BC classification. Consequently, the extracted feature vector was used as input to the PSO model's feature selection, thereby enabling significant feature selection. Thus, the outputs of the 6B-Net selection vector and ResNet-50 selection vector were sequentially combined to improve BC classification. However, the introduced approach required analyzing the integration of different feature selection methods to further enhance classification accuracy.

Yiping Zhou et al. [19] introduced a resolution-adaptive network for classifying breast cancer from histopathological images. Specifically, a new diagnostic technique was emerging using image processing and was widely accepted among pathologists. This research presented Anomaly Detection with SVM (ADSVM) and the Resolution Adaptive Network (RANet) for

https://doi.org/10.31436/iiumej.v27i2.3387

anomaly detection and adaptive image resolution. In the ADSVM method, a semi-supervised approach separates misleading and undesirable trained connections to improve the RANet model's training performance. However, to enhance classification performance and computational efficiency, a feature-extraction approach was required to reproduce the RANet model's features.

Mengyun Qiao et al. [20] developed a classification of BC on MRI-US by separating the modality-specific features using MUM-Net. This method effectively leverages paired multi-modality information to enhance breast tumor classification performance by providing complementary information from MRI and US images. The method currently focuses on joint predictions across dual modalities; however, once data for a single modality are available, this could limit its applicability in certain scenarios.

From the overall analysis, the existing approach has faced challenges such as a lack of accurate identification of breast tumors due to the localization of lesions and tumor prediction. Thus, in this research, the ensemble learning approach uses RNN, CNN, and LSTM networks, combined to make predictions using a soft voting mechanism.

3. PROPOSED METHODOLOGY

Specifically, 2D and 3D convolutional layers are separately applied to extract features from different modalities. Then, a discrimination-adaptation module is designed to learn modality-agnostic features via adversarial learning with three discriminators. Thereafter, a feature fusion module is designed to integrate modality-agnostic features and is made more compact using AM with NNS. Lastly, 3 different layers are set up for the fused features to obtain the joint prediction of expression levels. Figure 1 shows the workflow of the proposed model.

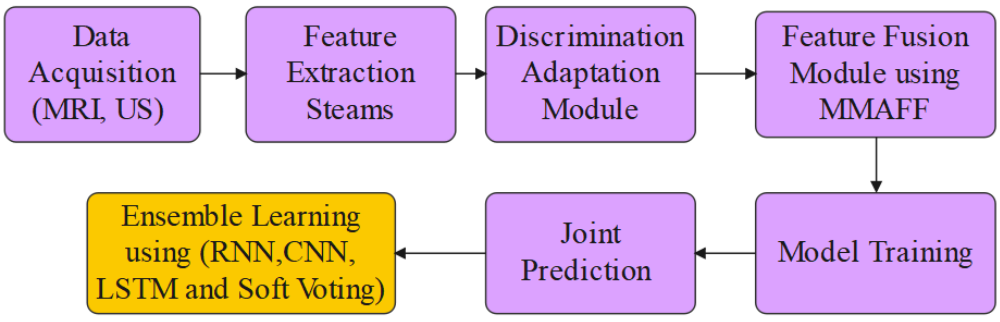


Figure 1. Workflow of the proposed model

3.1. Dataset Acquisition

This approach enrolled 502 patients who received breast US and MRI at the Cancer Center between 2017 and 2020. Institutional review board approval was acquired; however, informed consent was waived for all patients. Among the 502 patients, both MRI and ultrasound images were available for each case. In particular, the dataset comprises both benign and malignant tumors, with balanced representation. All images were reviewed and labeled by skilled radiologists based on pathological confirmation, and clinical indicators were thereby used as

auxiliary labels [21]. For the US examination, all images were acquired by a 7-15 MHz transducer using both the Aixplorer and the Resona 5S systems. A standardized imaging protocol was used: starting with the major axis, 12 2D images were captured with a 180-degree clockwise rotation between each image at equal intervals.

The multi-angle images could provide more comprehensive architectural data than conventional single-view US images. MR imaging in the prone position was performed on multiple MR scanners with breast coils, including a 3-T MR, a 1.5-T MR, and a 3-T MR. The contrast agent was injected intravenously through a compression nozzle at an injection rate of 1.5-3 mL/s, followed by a 15 mL flush with 0.9% NaCl at a similar injection rate. All patients underwent DCE-MRI, which included one pre-contrast image and four post-contrast images for each patient. Clinical features, including LNM, histological grade, and Ki-67, were collected from the medical system for all patients. The dataset of 502 cases was divided into training, validation, and testing sets in a 70:15:15 ratio, yielding 351 training samples, 75 validation samples, and 76 testing samples. All reported performance results are obtained by comparing the predicted outputs with the ground-truth labels of the independent test set.

3.2. Feature extraction based on multi-stream

A multi-stream feature extractor is used to extract feature embeddings from both modalities. After data acquisition, feature extraction is performed using 12 US images and 5 MRI sequences, each with a different convolutional kernel. Features are extracted from the left and right breasts using an MRI scanner, then concatenated and passed through the network's feature discrimination-adaptation and fusion steps to reduce complexity. The streams for each modality are separated in the shallow convolutional layers. The US data uses a ResNet-50V2 network, which adopts the well-known residual neural network design and has been shown to achieve superior performance. The MRI feature extractor uses a 3D ResNet-50V2 architecture with 3 × 3 × 3 convolutional kernels, and the features are subsequently reduced to 512 dimensions in the last convolutional layer. ResNet-50V2, which uses many layers, skip connections, and pre-trained weights, and has a high-accuracy profile, is a solid choice for feature extraction in breast tumor classification.

3.3. Discrimination-Adaptation Module

The discrimination–adaptation module is formulated to explicitly distinguish Modality-Specific Features (MSF) and Modality-Agnostic Features (MAF). Therefore, two discriminators are incorporated to classify the extracted features as belonging to the MRI or US modality. Additionally, during adversarial training, the feature extractor learns to confuse these discriminators, which facilitates the network to provide modality-agnostic representations while retaining modality-specific characteristics for complementary learning. This module is adopted by including a dual-modality discrimination and a single-adaptation module, with embedded feature loss and adversarial training. The two losses, i.e., L_M and L_U are imposed originally through the discrimination modules of the two modalities to achieve feature disentanglement: on the one hand, to separate the modality-specific features; on the other hand, to project the Modality-Specific features (MSF) to other patients' Modality-Agnostic features (MAF) for the same patient. The prediction error is formulated by Eqs. (1) and (2):

$$\mathrm{L_M} = \mathbb{E}_{\mathrm{M}}\left[\left\|\mathrm{g_{w_m}}(\mathrm{F_M^s}) - \mathrm{F_M^\alpha}\right\|\right] \quad (1)$$

https://doi.org/10.31436/iiumej.v27i2.3387

$$L_U = \mathbb{E}_U\left[\left\|g_{w_U}(F_U^s) - F_U^\alpha\right\|\right] \quad (2)$$

where $g\omega M$, $g\omega U$ signify discriminators for MRI and US, respectively. In the generation, the MSF is uncorrelated with MAF to mislead the prediction. Alternatively, the feature extractor learns MSF from MAF by adjusting the prediction loss. At that time, MSF had 3 fully connected layers, and to eliminate the MAF, which is expressed by Eq. (3);

$$L_{mo} = \mathbb{E}_m\left[-\log\left(p\left(y_m|F_m^a, g_{w_a}\right)\right)\right] \quad (3)$$

where g_{w_a} characterizes the modality adaptation constraints. $p\left(y_m|F_m^a, g_{w_a}\right)$ refers to the predicted probability of F_m^a which belongs to the ground-truth modality y_m which is a one-hot-encoded label equivalent to $\{0, 1\}$. This module confirms that the features from MAF and MSF are autonomous and encompass less unnecessary data.

3.4. Feature Fusion Module

The MMAFF framework fuses modality-agnostic features by modeling both spatial relationships and temporal dependencies. Specifically, this operation combines Spatial Attention Graph Convolution (SAGC), which helps capture spatial correlations, and Temporal Adaptive Feature Fusion (TAFF), which dynamically aggregates temporal features, thereby ensuring robust fusion across heterogeneous modalities. Thus, the modality-agnostic features obtained from the discrimination–adaptation module are embedded into the MMAFF framework for structured spatial and temporal fusion.

3.4.1. Multi-Modality Adaptive Feature Fusion Graph Convolutional Network

The Global Average Grouping (GAP) and fully connected layers are used to convert the outputs from the previous layers, yielding layer-by-layer predictions for distinct streams. MMAFF comprises SAGC and TAFF. SAGC is used to obtain data from spatial dimensions, while TAFF is used to extract temporal features, along with the initial features.

3.4.2. SAGC Module

This approach adopts a dynamic setting for creating a spatial attention map convolution. The model simultaneously feeds initial information into 2 equivalent divisions, σ, φ and ψ that are managed through 1×1 convolution operation and a Temporal Pooling (TP) block. This feature map is added through the adjacency matrix A_p for attaining optimal A_{cwt}. This module performs attentional feature fusion to replace the actual residual integration used to combine spatial and temporal-scale data, as expressed in Eq. (4).

$$A_{cwt} = \beta Q(X_{in}) + A_p \quad (4)$$

β refers to learnable constraints; A_p refers to the shared topology at $p - th$ channel; Q denotes a topological relationship of the definite channel, which is demarcated as shown in Eq. (5);

$$Q(X_i) = \sigma\left(TP\left(\phi(X_{in})\right) - TP\left(\psi(X_{in})\right)\right) \quad (5)$$

https://doi.org/10.31436/iiumej.v27i2.3387

This approach attains the channel-wise model A_{cwt}; Then, the initial features of 1×1 convolution multiply the outcomes with A_{cwt} to combine the spatial-dimension data, which is mentioned in Eq. (6):

$$X_{out} = A_{cwt} \otimes \left(\theta(X_{in})\right) \tag{6}$$

θ refers to 1 × 1 convolution; the operation of matrix multiplication is referred to as $\otimes$.

3.4.3. TAFF Module

The multi-scale temporal adaptive feature fusion comprises two components: Temporal Adaptive (TA) and Temporal Feature Fusion (TFF). TA adjusts the convolution kernel size and dilation rate across several layers using a robust attention mechanism. This part is enhanced by the multi-scale temporal convolution, which comprises 4 branches that use 1 × 1 convolution to reduce the channel dimension. This module exploits the formulation, which is stated as Eq. (7):

$$t = \frac{abs(log\, C_l,2)+b}{gamma} \tag{7}$$

C_l refers to the output channel dimension at $l - th$ network layer; $gamma$ and b stated as mapping-function constraints, varied to 2 and 1, respectively. The four branches at various scales are obtained by the aggregation function. This module does not introduce additional branches; consequently, no modification to the constraint or computational complexity. In another part of this module, an attention feature fusion combines contextual data of various dimensions with the channel. This module exploits AFF to combine features of various branches and exploits dual-input branches X and X_1. But this module focuses on the initial feature data and the temporal dimension to enable efficient feature fusion. The combination of initial and temporal features addresses the context aggregation issue, enhancing modeling efficiency, as stated in Eq. (8).

$$X' = X \otimes M(\cdot) + X_1 \otimes \left(1 - M(\cdot)\right) \tag{8}$$

The input residual connection is referred to as X; X_1 refers to the aggregated output obtained by multi-scale convolution. The formulation for $M(\cdot)$ is stated in Eq. (9);

$$M(\cdot) = \left(Sigmoid\left(L(X \uplus X_1) \oplus G(X \uplus X_1)\right)\right. \tag{9}$$

$L(\cdot)$ & $G(\cdot)$ refer to the local and global channel context, correspondingly. In the attention module, the data from $L(\cdot)$ are combined with $G(\cdot)$ data The primary fusion is accomplished on X and . After applying the sigmoid, the output is displayed from 0 to 1. This approach takes the weighted average and subtracts the fusion weights 1, which is referred to as soft selection [22]. The network model calculates its weights during training.

3.5. Joint Prediction

The MRI, US, and combined MAF are summarized as a final set and processed into 3 distinct layers for joint prediction of indicators. The final task loss integrates 3 losses, which are expressed by Eqs. (10) to (13);

$$\mathrm{L_{cls}} = \mathrm{L_{t_1}} + \mathrm{L_{t_2}} + \mathrm{L_{t_3}} \tag{10}$$

https://doi.org/10.31436/iiumej.v27i2.3387

$$L_{t_1} = \mathbb{E}_m\left[-log\left(p\left(y_{t_1}|[F^a, F_M, F_U], f_{\theta_{t_1}}\right)\right)\right] \quad (11)$$

$$L_{t_2} = \mathbb{E}_m\left[-log\left(p\left(y_{t_2}|[F^a, F_M, F_U], f_{\theta_{t_2}}\right)\right)\right] \quad (12)$$

$$L_{t_3} = \mathbb{E}_m\left[-log\left(p\left(y_{t_3}|[F^a, F_M, F_U], f_{\theta_{t_3}}\right)\right)\right] \quad (13)$$

F_M and F_U signifies the MAF and MSF sets in MRI and US. t_1, t_2 and t_3 denote the classification task of LNM, HG, and Ki-67.

3.6. Model Training

The total loss involves the feature loss, including , L_U and L_C for decoupling the features, task loss, including , L_{t_2}, and for predicting the three indicators. The min-max is implemented in the network training process and used by Eqs. (14) and (15);

$$L_f = L_{cls} + L_c - \lambda_1 L_{mo} - \lambda_2(L_M + L_U) \quad (14)$$

$$L_g = -\lambda_1 L_{mo} - \lambda_2(L_M + L_U) \quad (15)$$

where λ_1 and λ_2 are fixed at 0.2. There are 2 steps involved: first, minimize, which increases L_{mo} and the classification losses, L_{fU}, then increase L_g in various epochs. The alternative learning process is expressed by Eqs. (16) and (17):

$$f_\theta = arg\min_\theta L_f \quad (16)$$

$$g_{w_a}, g_{w_M}, g_{w_U} = arg\max_{w_a, w_M, W_U} L_g \quad (17)$$

where, g_{ω_N} signifies the overall network constraints.

4. CLASSIFICATION

Ensemble learning comprises the parallel training of multiple classifiers to improve performance in breast tumor analysis. Deep learning (DL) techniques are implemented to combine the training experiences of CNN, RNN, and LSTM for further processing. Here, three classifiers are considered in the ensemble model to produce efficient results. A proposed RNN model for BC classification consists of an input layer, 5 hidden layers, and 5 dropout layers [23]. The input layer contains an equal number of features for neurons, ReLU activation functions, and uniform kernel initialization. LSTMs differ from RNNs in how they compute the hidden state, making them more effective for sequential data such as patient history or time-series data related to tumor growth.

The LSTM [23] network is an improved version of the RNN that can learn dependencies for sequence prediction. The LSTM model incorporates enhancements to the RNN design to implement a suitable backpropagation (BP) training method. The LSTM model addresses the vanishing gradient problem prevalent in standard RNNs. This issue arises when gradients diminish over time due to repeated weight updates, but an LSTM mitigates it by propagating errors through its recurrent connections. CNN is a feedforward neural network composed of six layers: an input layer, a convolutional layer, a non-linear layer, a pooling layer, a fully

https://doi.org/10.31436/iiumej.v27i2.3387

connected layer, and an output layer. To train a CNN, the mapping function is determined using the feedforward operation, and the loss function is optimized using propagation techniques, particularly the gradient descent algorithm. The main aim of DL is to achieve multiple levels of feature classification. CNNs are popular in image classification due to their ability to detect patterns and edges. A CNN model trained on image data captures spatial features, making it particularly effective for image classification tasks. The final class label is determined by the class with the highest total probability as expressed in Eq. (18):

$$\hat{y}_i = arg\,max_k \sum_{j=1}^{m} w_j p_{i,k}^{j} \tag{18}$$

where, m represents the classifier count, weight of j^{th} classifier is stated as w_j; $p_{i,}^{j}$ denoted the prediction probability of j^{th} classifier for passing i^{th} sample to k^{th} classes.

5. Experimental Results and Discussion

In this research, the MRI and US images were regularized to [0, 1] and resized to 5 × 256 × 256 × 48 and 12 × 256 × 256, respectively. This study implemented the ensemble classifiers with PyTorch and a graphics processing unit (NVIDIA Tesla V100 32 GB). Assessment parameters, such as accuracy, precision, sensitivity, Area Under Curve (AUC), and specificity, are used to estimate model performance. The AUC is evaluated from the Receiver Operating Characteristic (ROC) curve, which plots the true positive rate against the false positive rate at various classification thresholds. Specifically, the AUC is the area under the ROC curve, which measures the probability that the classifier ranks a randomly selected positive sample higher than a randomly selected negative sample. The mathematical formula is given in Equation. (19-22),

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{19}$$

$$Precision = \frac{TP}{TP+FP} \tag{20}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{21}$$

$$Specificity = \frac{TN}{TN+FP} \tag{22}$$

where, TP, TN denotes True Positive and True Negative; FP and FN specifies False Positive and False Negative correspondingly.

5.1. Performance Analysis

The performance of the proposed method is evaluated on the collected MRI-US dataset. ResNet-50-v2 is used to extract essential features, while MMAFF processes these features to analyze temporal relationships in the data. The proposed approach achieved better performance across all indicators, including LNM, HG, and Ki-67 levels. Table 1 and Figure 2 represent the performance analysis of clinical indicators; Table 2 and Figure 3 represent the performance analysis of classification on the MRI-US dataset.

https://doi.org/10.31436/iiumej.v27i2.3387

Table 1. Performance analysis of clinical indicators

Methods	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	AUC (%)
Lymph Node Metasis	82.5	82.7	81.9	81.8	85.7
Histological Grade	85.7	85.4	84.8	84.6	87.3
Ki-67 level	89.6	89.7	88.6	88.9	91.2

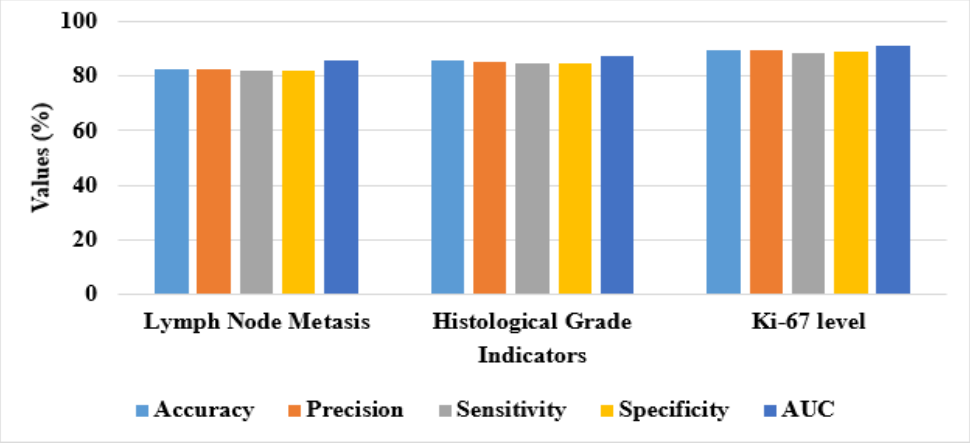


Figure 2. Graphical illustration of the clinical indicator results.

Table 2. Performance analysis of the classification on the MRI-US dataset

Methods	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	AUC (%)
CNN	96.85	83.75	96.23	92.75	84.15
LSTM	97.15	83.96	97.45	93.25	85.26
RNN	98.52	84.16	98.25	94.63	86.12
Ensemble Learning	99.25	85.45	99.45	95.99	87.45

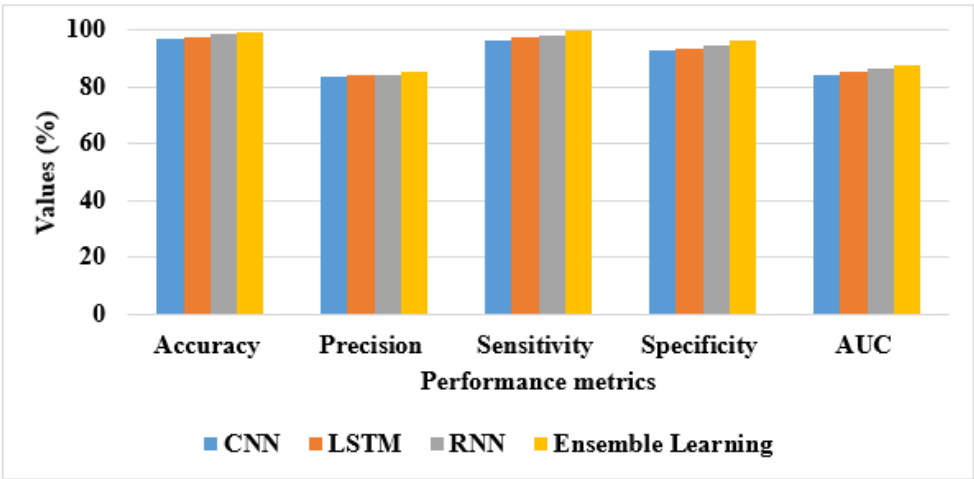


Figure 3. Graphical illustration of classification results.

From Table 2, it is observed that existing methods employing feature selection techniques, such as CNN, LSTM, and RNN, were evaluated. The ensemble learning method achieves a

high accuracy of 99.25%, a precision of 97.62%, a recall of 96.02%, and an F1-score of 96.79%.

5.2. Comparative Analysis

The proposed model is compared with existing approaches, including BPBRW-HKH-ABO [16], ADSVM [19], and MUM-Net [20], as shown in Table 3 and Figure 4. Accuracy, precision, sensitivity, and specificity are used as evaluation metrics to assess model performance.

Table 3. Comparison of the proposed model

Methods	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	AUC (%)
BPBRW-HKH-ABO [16]	99.06	N/A	99.09	95.78	N/A
RANet [19]	82.4	82.9	82.1	82.3	84.1
MUM-Net [20]	85.4	85.6	84.9	85.4	87.8
Proposed Ensemble Learning Method	99.25	85.45	99.45	95.99	87.45

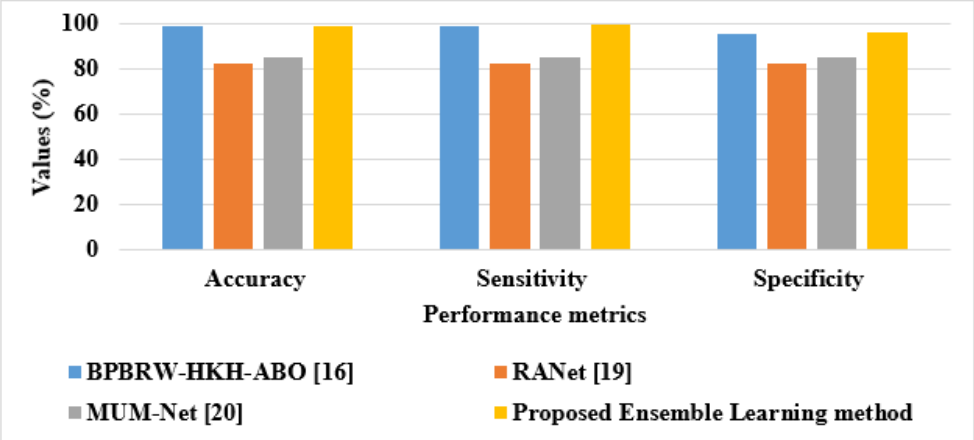


Figure 4. Graphical illustration of comparative analysis results

From Table 3 and Figure 4, it is observed that, compared with various existing models, the proposed approach achieved accuracies of 82.5%, 85.7%, and 89.6% for LNM, HG, and Ki-67, respectively.

5.3. Discussion

The proposed ensemble learning method is introduced to enhance patient survival in BC and thereby enable early and accurate diagnosis. In particular, the proposed model uses a feature fusion module to integrate modality-agnostic features. Thereby, it is necessary to distinguish between two different evaluation levels in this research. Specifically, the accuracy of 89.6% indicates that the clinical indicator prediction (Ki-67 level) incorporating the fused MMAFF features is accurate, as demonstrated in Table 1. Further, the accuracy of 99.25% pertains to the final breast tumor classification, which uses a soft-voting ensemble of CNN, RNN, and LSTM on the MRI-US dataset, as illustrated in Table 2. Hence, these values signify

different tasks and are therefore not contradictory. The ensemble model outperforms individual CNN, RNN, and LSTM models because each model captures complementary characteristics: CNNs assist in extracting spatial image features, RNNs capture sequential patterns, and LSTMs handle long-term dependencies. Thus, soft voting combines these heterogeneous predictions, thereby reducing model bias and improving generalization. Despite strong performance, the research has limitations: the dataset comprises 502 cases collected from a single medical center, which is a moderate sample size for deep learning–based medical image analysis and may limit generalizability. Thus, future work should validate the framework on multi-center datasets and include additional imaging modalities.

6. CONCLUSION

In the proposed approach, an ensemble deep learning method is proposed for breast tumor classification from MRI and US. In this research, the ensemble learning approach uses RNN, CNN, and LSTM networks, combined via soft voting to make predictions. First, the features are extracted from 2D and 3D convolutional layers. Then, a feature fusion module is proposed that uses MMAFF to maximize feature compactness. The MRI-US dataset comprises 502 cases and is used to assess the developed ensemble learning approach. Compared with existing techniques such as Deep Convolutional Neural Networks (DCNN) and Resolution Adaptive Network (RANet), the proposed ensemble learning technique achieves 89.6% accuracy for clinical indicator prediction and 99.25% accuracy for final tumor classification on the MRI-US dataset. In future work, novel optimization techniques can be used to solve high-dimensional optimization problems and consider more classes.

REFERENCES

[1] Zou Y, Zhang J, Huang S, Liu B. (2022) Breast cancer histopathological image classification using attention high-order deep network. International Journal of Imaging Systems and Technology, 32(1):266-279.

[2] Liu M, He Y, Wu M, Zeng C. (2022) Breast histopathological image classification method based on autoencoder and siamese framework. Information, 13(3):107.

[3] Ahmad N, Asghar S, Gillani SA. (2022). Transfer learning-assisted multi-resolution breast cancer histopathological images classification. The Visual Computer, 38(8):2751-2770.

[4] Alqahtani Y, Mandawkar U, Sharma A, Hasan MNS, Kulkarni MH, Sugumar R. (2022) Breast cancer pathological image classification based on the multiscale CNN squeeze model. Computational Intelligence and Neuroscience, 2022(1):7075408.

[5] Zhou Y, Zhang C, Gao S. (2022) Breast cancer classification from histopathological images using resolution adaptive network. IEEE Access, 10:35977-35991.

[6] Michael E, Ma H, Li H, Qi S. (2022) An optimized framework for breast cancer classification using machine learning. BioMed Research International, 2022(1):8482022.

[7] Wakili MA, Shehu HA, Sharif MH, Sharif MHU, Umar A, Kusetogullari H, Ince IF, Uyaver S. (2022) Classification of breast cancer histopathological images using DenseNet and transfer learning. Computational Intelligence and Neuroscience, 2022(1):8904768

[8] Balkenende L, Teuwen J, Mann RM. (2022) Application of deep learning in breast cancer imaging. In Seminars in Nuclear Medicine. WB Saunders, 52(5):584-596

[9] Liu M, Hu L, Tang Y, Wang C, He Y, Zeng C, Lin K, He Z, Huo W. (2022) A deep learning method for breast cancer classification in the pathology images. IEEE Journal of Biomedical and Health Informatics, 26(10):5025-5032.

[10] Rakha EA, Tse GM, Quinn CM. (2023) An update on the pathological classification of breast cancer. Histopathology, 82(1):5-16.

[11] Yu D, Lin J, Cao T, Chen Y, Li M, Zhang X. (2023) SECS: An effective CNN joint construction strategy for breast cancer histopathological image classification. Journal of King Saud University-Computer and Information Sciences, 35(2):810-820.

[12] Jabeen K, Khan MA, Balili J, Alhaisoni M, Almujally NA, Alrashidi H, Tariq U, Cha, JH. (2023) BC2NetRF: breast cancer classification from mammogram images using enhanced deep learning features and equilibrium-jaya controlled regula falsi-based features selection. Diagnostics, 13(7):1238.

[13] Shankar K, Dutta AK, Kumar S, Joshi GP, Doo IC. (2022) Chaotic sparrow search algorithm with deep transfer learning enabled breast cancer classification on histopathological images. Cancers, 14(11):2770.

[14] Hamedani-KarAzmoudehFar F, Tavakkoli-Moghaddam R, Tajally AR, Aria SS. (2023) Breast cancer classification by a new approach to assessing deep neural network-based uncertainty quantification methods. Biomedical Signal Processing and Control, 79:104057.

[15] Rashmi R, Prasad K, Udupa CBK. (2023) Region-based feature enhancement using channel-wise attention for classification of breast histopathological images. Neural Computing and Applications, 35(8):5839-5854.

[16] Ahmad N, Asghar S, Gillani SA. (2022) Transfer learning-assisted multi-resolution breast cancer histopathological images classification. The Visual Computer, 38(8):2751-2770.

[17] Ukwuoma CC, Hossain MA, Jackson JK, Nneji GU, Monday HN, Qin Z. (2022) Multi-classification of breast cancer lesions in histopathological images using DEEP_Pachi: Multiple self-attention head. Diagnostics, 12(5):1152.

[18] Clement D, Agu E, Suleiman MA, Obayemi J, Adeshina S, Soboyejo W. (2022) Multi-class breast cancer histopathological image classification using multi-scale pooled image feature representation (MPIFR) and one-versus-one support vector machines. Applied Sciences, 13(1):156.

[19] Umer MJ, Sharif M, Kadry S, Alharbi A. (2022) Multi-class classification of breast cancer using 6B-Net with deep feature fusion and selection method. Journal of Personalized Medicine, 12(5):683.

[20] Zhou Y, Zhang C, Gao S. (2022) Breast cancer classification from histopathological images using resolution adaptive network. IEEE Access, 10:35977-35991.

[21] Qiao M, Liu C, Li Z, Zhou J, Xiao Q, Zhou S, Chang C, Gu Y, Guo Y, Wang Y. (2022) Breast tumor classification based on MRI-US images by disentangling modality features. IEEE Journal of Biomedical and Health Informatics, 26(7):3059-3067.

[22] Akyol K, Uçar E, Atila Ü, Uçar M. (2024) An ensemble approach for classification of tympanic membrane conditions using soft voting classifier. Multimedia Tools and Applications, 1-22.

[23] Saleh H, Abd-El Ghany SF, Alyami H, Alosaimi W. (2022) Predicting breast cancer based on optimized deep learning approach. Computational Intelligence and Neuroscience, 2022(1):1820777.

https://doi.org/10.31436/iiumej.v27i2.3683

New Control Strategy of Battery Energy Storage System in Microgrids Integrated with Wind and Solar Power Plants

DUY QUANG NGUYEN[1,2], MINH KHOA NGO[2*]

[1]*Vinh Son - Song Hinh Hydropower Joint Stock Company, Gia Lai province, Vietnam*
[2]*Faculty of Engineering and Technology, Quy Nhon University, Gia Lai province, Vietnam*

**Corresponding author: ngominhkhoa@qnu.edu.vn*

(Received: 8 April 2025; Accepted: 22 January 2026; Published online: 10 May 2026)

ABSTRACT: The global shift toward renewable energy, particularly wind and solar, has led to the rapid growth and diversification of distributed energy resources (DERs). However, integrating these sources sustainably and efficiently presents major challenges for researchers, investors, and power system operators. Their impact on distribution grids, along with operational uncertainties and disturbances, requires a comprehensive, system-wide perspective. This study proposes targeted strategies to enhance renewable energy utilization, offering both economic and environmental benefits while maintaining technical reliability and energy security. The study develops a microgrid model that includes a solar plant, a wind farm, an energy storage system, a distribution grid, and loads. As the core component, the battery energy storage system (BESS) plays a vital role in grid stabilization, power quality enhancement, and voltage/frequency regulation through advanced control strategies. Simulation results are analyzed to assess grid performance, evaluate the influence of renewable sources, and verify the effectiveness of the BESS in maintaining stable microgrid operation under realistic operating conditions.

ABSTRAK: Peralihan global ke arah tenaga boleh diperbaharui, khususnya angin dan solar, telah membawa kepada pertumbuhan pesat serta kepelbagaian sumber tenaga teragih (DERs). Walau bagaimanapun, pengintegrasian sumber-sumber ini secara mampan dan cekap menimbulkan cabaran besar kepada para penyelidik, pelabur, dan pengendali sistem kuasa. Kesan sumber tersebut terhadap grid pengagihan, bersama-sama dengan ketidakpastian operasi dan gangguan, memerlukan perspektif menyeluruh pada skala sistem. Kajian ini mencadangkan strategi berfokus bagi meningkatkan penggunaan tenaga boleh diperbaharui, yang menawarkan manfaat ekonomi dan alam sekitar sambil mengekalkan kebolehpercayaan teknikal serta keselamatan tenaga. Kajian ini membangunkan model mikrogrid yang merangkumi loji solar, ladang angin, sistem penyimpanan tenaga, grid pengagihan, dan beban. Sebagai komponen teras, sistem penyimpanan tenaga bateri (BESS) memainkan peranan penting dalam penstabilan grid, peningkatan kualiti kuasa, serta pengawalan voltan/frekuensi melalui strategi kawalan lanjutan. Dapatan simulasi dianalisis bagi menilai prestasi grid, menilai pengaruh sumber boleh diperbaharui, dan mengesahkan keberkesanan BESS dalam mengekalkan operasi mikrogrid yang stabil pada keadaan operasi realistik.

KEYWORDS: *Battery energy storage system, Droop controller, Microgrid, Grid-following, Grid-forming.*

1. INTRODUCTION

Advances in power electronics, particularly in inverter technology, have driven major progress in electrical engineering and automation control. Renewable energy sources (RES),

such as wind and solar, are increasingly deployed to replace conventional energy sources [1, 2]. Recent market trends further emphasize this global transition. In 2024, global investment in the energy transition reached USD 2.1 trillion, including USD 728 billion in renewables and USD 54 billion in energy storage systems, marking an 11% increase compared to the previous year [3]. That same year, global renewable power capacity increased by 18%, reaching 741 GW, with solar and wind accounting for the majority of newly installed capacity, solar alone representing 81% and wind 16% of the total [4]. The grid-forming inverter market is also projected to grow from USD 680 million in 2023 to USD 1,042 million by 2028, at a compound annual growth rate (CAGR) of 8.9% [5]. In parallel, battery energy storage deployment is accelerating, with 12.3 GW added in the U.S. in 2024 alone, a 33% year-over-year increase, and global lithium-ion battery demand exceeded 1 TWh [6]. These statistics underscore the urgency of developing effective control strategies for inverter-based resources and battery energy storage systems, reinforcing the study's real-world relevance. However, this transition has introduced significant challenges to power systems, particularly in distribution grids, impacting their stability and reliability. This study investigates typical fault conditions and operational scenarios in distribution networks, including both grid-connected and islanded microgrid modes. A series of simulations is conducted on RES-based microgrids using inverter control strategies such as grid-following, grid-forming, and droop control [7-10].

A microgrid is a part of the distribution grid, functioning as an independent electrical network that enables on-site power generation, connection, monitoring, and control of distributed energy resources. This configuration improves efficiency, enhances grid stability, and increases operational resilience. In recent years, many studies have focused on modeling and controlling microgrids, as well as fault management and real-time operational modeling under the uncertainty of renewable energy generation [11, 12]. Various control methods for inverters operating in grid-forming mode have been developed and are currently in use worldwide, including droop control, virtual synchronous machine (VSM), virtual oscillation control (VOC), and virtual impedance control (VIC) [13, 14]. These methods enable grid-forming inverters to support the grid through various functions, such as operating in islanded mode, ensuring synchronization with other generators, increasing system inertia, supporting primary and secondary frequency control, providing reactive power, adjusting voltage, damping frequency/angle oscillations, and enabling black-start capability. Among these methods, droop control is the most widely used and is adopted in this study. It establishes a linear relationship between frequency (or voltage) and the inverter's active (or reactive) power output. As frequency (or voltage) decreases, the inverter proportionally increases its active (or reactive) power output, and vice versa. This mechanism allows multiple inverters to share the load based on their rated capacities.

In microgrids, Battery Energy Storage Systems (BESS) play a critical role in stabilizing the grid and enhancing operational flexibility through controlled charging and discharging processes [15, 16]. Through power electronic inverters, BESS can independently and simultaneously regulate both active and reactive power, enabling functions including frequency support, voltage regulation, and power smoothing. As a result, BESS-based inverters are increasingly replacing traditional reactive power compensation devices due to their faster response and greater controllability [17].

This paper investigates the integration of wind power, solar power, and battery energy storage systems (BESS) in a distribution-level microgrid through mathematical modeling and real-world data in the MATLAB/Simulink environment. The study evaluates the role of BESS in enhancing microgrid stability and operational efficiency under both grid-connected and islanded modes. The main contributions of this paper are as follows: (i) Developing dynamic

models and control strategies for wind, solar, and BESS subsystems connected to the distribution grid; (ii) Designing representative operational scenarios and fault conditions to assess microgrid performance under various disturbances; (iii) Conducting simulation-based analysis and evaluation of technical performance metrics, along with recommendations to improve microgrid operation.

Table 1. Summarizes representative studies on inverter-based microgrids and highlights the key differences and novel contributions of this paper

Study/ Ref.	Microgrid Mode	RES Considered	BESS Role	Control Strategy	Fault Scenarios	Contribution Summary
Anttila et al. (2022) [18]	Grid-connected & Islanded	General DERs	Not explicitly included	Droop, VSM, VOC, VIC (review)	✗	Review of grid-forming inverter control methods without new simulation or RES-BESS integration
Lamichhane & Adhikari (2023) [19]	Islanded	Solar + V2G	Inertia emulation via virtual synchronous machine	VSG control for frequency stability	✗	Focused on VSG-based frequency control in solar-based microgrid; lacks transient and BESS-aware studies
Elwakil et al. (2023) [20]	Islanded	PV-only	PV-inverter support	Bang-bang VSG control	✗	Fast frequency response via VSG, but no multi-source modeling or BESS involvement
Chakraborty et al. (2022) [21]	Islanded	Wind + PV	Implicit support via control strategy	μ-synthesis for robust GFL/GFM	✗	Hardware-in-loop validated robust controllers under wind and solar; no dedicated fault-ride-through modeling
Lin et al. (2022) [22]	Grid-connected	General RES	Voltage regulation via grid-forming inverter	Model-free predictive voltage control (MFPC)	✗	Improved voltage stability in low-inertia systems; lacks integrated fault and performance evaluation
This work	Grid-connected & Islanded	Solar + Wind + BESS	Full dynamic grid-forming support (frequency & voltage stabilization)	GFL, GFM + droop control with BESS PQ regulation	✓	Integrates detailed RES+BESS dynamic models; simulates realistic faults, islanding, and reconnection scenarios using MATLAB/Simulink

The remainder of this paper is organized as follows: Section 1 outlines the current state of the distribution grid, control strategies for renewable energy systems, related work, and the research objectives. Section 2 presents the mathematical models of the distribution grid, distributed generators (DGs), BESS, and microgrid components, along with simulation models developed in MATLAB/Simulink and representative operational scenarios. Section 3 presents

the simulation results and evaluates the system's performance under different operating conditions. Finally, Section 4 provides the conclusions.

2. Proposed System and Method

2.1. DFIG-based wind turbine modeling

This study investigates and integrates a Type-3 wind turbine model, based on a Double-Fed Induction Generator (DFIG), into the microgrid simulation framework. The DFIG GE 1.5 wind turbine model, as shown in Figure 1, was studied and developed by Nicholas W. Miller and colleagues at GE Power Systems. It has been implemented in Simulink and used in simulations by researchers at IREQ, Hydro-Québec, Varennes, QC, Canada [23].

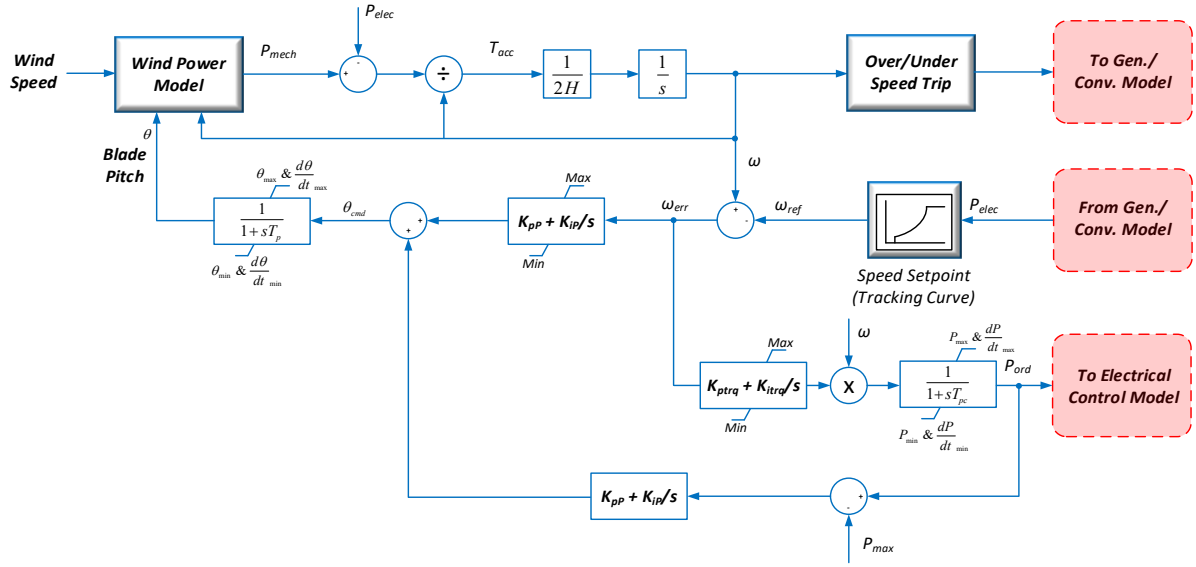


Figure 1. Dynamic modeling of a DFIG-based wind turbine.

The aerodynamic torque generated by the wind turbine is calculated as:

$$T = \frac{1}{2} C_p(\lambda, \beta) \pi \rho R^5 \frac{v_\omega^3}{\lambda^3} \tag{1}$$

where ρ is the air density (kg/m³), R is the blade radius (m), v_ω is the wind speed (m/s), β is the blade pitch angle (degrees), and λ is the tip-speed ratio (ratio of blade tip speed to wind speed).

The power coefficient C_p, representing the wind-to-mechanical energy conversion efficiency, is given by:

$$C_p(\lambda, \beta) = 0.5176 \left(\frac{116}{\lambda_i} - 0.4\beta - 5 \right) e^{\frac{-21}{\lambda_i}} + 0.0068\lambda \tag{2}$$

where:

$$\frac{1}{\lambda_i} = \frac{1}{0.08\beta + \lambda} - \frac{0.035}{\beta^3 + 1} \tag{3}$$

The AC/DC/AC converter is divided into two main components: the rotor-side converter (C_{rotor}) and the grid-side converter (C_{grid}). Both C_{rotor} and C_{grid} are voltage source converters (VSCs) employing insulated-gate bipolar transistors (IGBTs) with forced switching to convert DC voltage into AC voltage. A capacitor, C, is connected across the DC-link, serving as a DC

https://doi.org/10.31436/iiumej.v27i2.3683

voltage source. Additionally, an inductor, L, is used to connect C_{grid} to the power grid [24], [25]. The three-phase rotor winding is connected to C_{rotor} via slip rings and brushes, while the stator winding is directly connected to the grid. The mechanical power captured by the wind turbine is converted into electrical power by the induction generator and delivered to the grid through the rotor and stator windings. The control system governs the wind turbine operation by generating commands for adjusting the blade pitch angle and producing control signals for the voltages V_{rc} for C_{rotor} and V_{gc} for C_{grid} to regulate the wind turbine's power, DC-link voltage, reactive power, and output voltage at the grid connection point.

In general, the absolute value of the slip coefficient s is much smaller than 1, and thus the rotor-side power P_r is significantly smaller than the stator-side power P_s. Since the mechanical torque on the rotor shaft T_m is positive for power generation, and ω_s is positive and constant due to the fixed grid frequency, the sign of P_r is a function of the slip coefficient s. P_r is positive for negative s values (i.e., when the rotor speed exceeds synchronous speed) and negative for positive s values (*i.e., when the rotor speed is below synchronous speed*). For super-synchronous operation, P_r is transmitted to the DC-link capacitor, increasing the DC voltage. For sub-synchronous operation, P_r is drawn from the DC-link capacitor and tends to decrease the DC voltage. The grid-side converter (C_{grid}) is used to either generate or absorb power (P_{gc}) to maintain a constant DC voltage. In the steady state, for a lossless AC/DC/AC converter, P_{gc} equals P_r, and the wind turbine speed is determined by the power P_r absorbed or generated by C_{rotor}. The phase sequence of the AC voltage generated by C_{rotor} is positive at sub-synchronous speeds and negative at super-synchronous speeds. The frequency of this voltage is equal to the product of the grid frequency and the absolute value of the slip. Both C_{rotor} and C_{grid} are capable of generating or absorbing reactive power and can be used to control reactive power or voltage at the grid connection point. To ensure that the DFIG operates as a generator in both modes, the power converters on both sides - the generator side and the grid side, as shown in Figure 2, must be inverters that support bidirectional power flow control. The converters enable the DFIG to operate in all four quadrants of the $P-Q$ complex plane, meaning the DFIG can supply reactive power to the grid (unlike conventional induction machines). The reactive power exchanged between the DFIG and the grid can be controlled independently of the active power [26, 27].

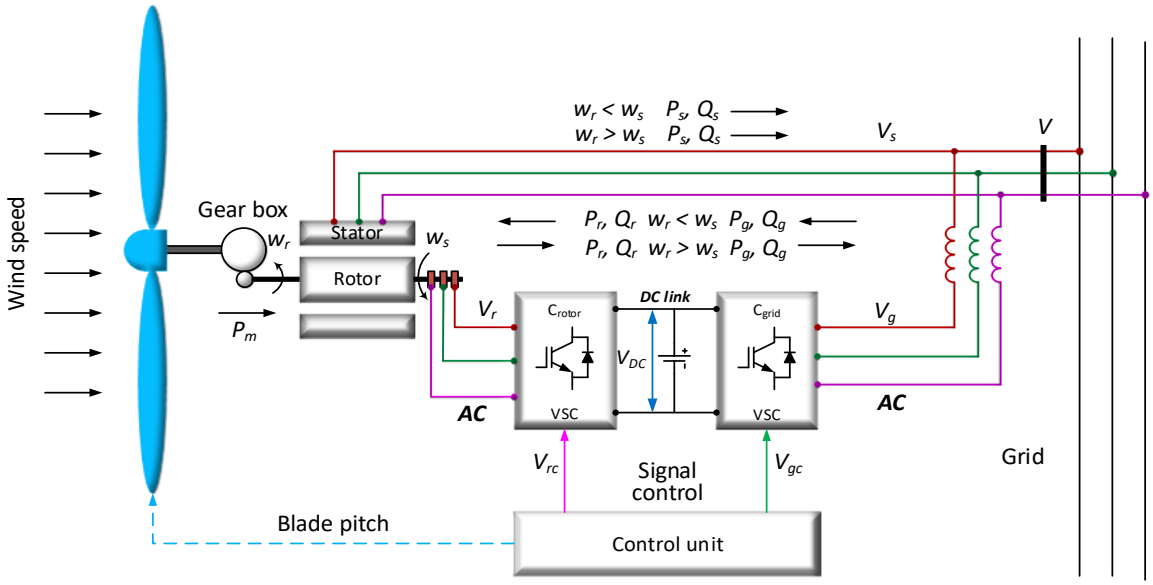


Figure 2. Control system of DFIG.

https://doi.org/10.31436/iiumej.v27i2.3683

The C_{rotor} converter is used to control the wind turbine's output power and the voltage (or the reactive power) measured at the grid terminals. The actual electrical output power, measured at the wind turbine's grid terminals, is combined with the total power losses (both mechanical and electrical) and compared to the reference power obtained from the MPPT tracking characteristic. A proportional-integral (PI) controller is used to reduce the power error to zero. The output of this controller is the reference rotor current I_{qr_ref}, which must be supplied to the rotor by the C_{rotor} converter. This is the current component responsible for generating the electromagnetic torque T_{em}. The actual I_{qr} component of the positive-sequence current is compared to I_{qr_ref}, and the PI current regulator minimizes the error. The output of the current controller is the voltage V_{qr} generated by the C_{rotor}. The current controller is supported by feedforward compensation terms to predict V_{qr}. The voltage or reactive power at the grid terminals is controlled by the reactive current flowing through the C_{rotor} converter, as shown in Figure 3.

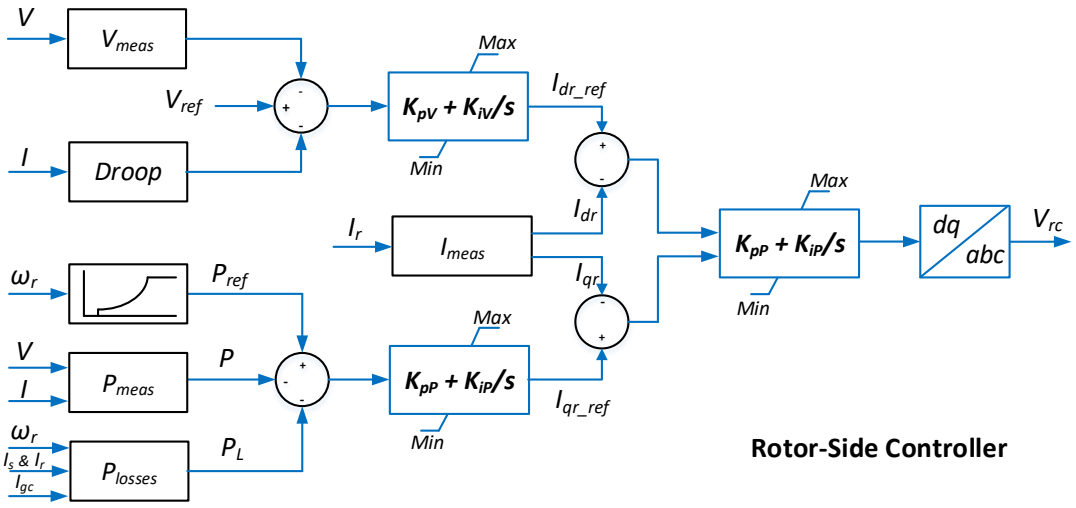


Figure 3. Rotor-side controller of DFIG.

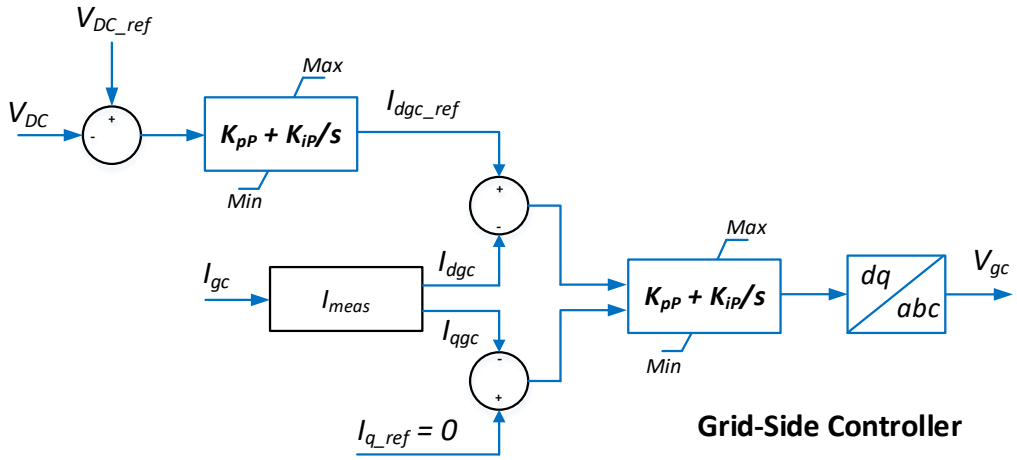


Figure 4. Grid-side controller of DFIG.

When the wind turbine operates in VAR (Volt-Ampere Reactive) control mode, the VAR controller maintains a constant reactive power at the grid terminals. The output of the voltage regulator or VAR controller is the d-axis reference current I_{dr_ref}, which must be supplied to

the rotor by the C_{rotor} converter. The current regulator, similar to the power controller, is used to regulate the actual I_{dr} component of the positive-sequence current to match its reference value. The output of this regulator is the d-axis voltage V_{dr} generated by the C_{rotor} converter. The current regulator is supported by feedforward compensation terms that predict V_{dr}. The C_{grid} converter is used to regulate the DC-link capacitor voltage. Additionally, this model allows the C_{grid} converter to generate or absorb reactive power, as illustrated in Figure 4.

2.2. Solar power system modeling

Solar energy is a form of energy harnessed from sunlight and converted into electricity through photovoltaic (PV) cells. Solar power is a clean, renewable energy source that has become increasingly popular in recent years due to its numerous benefits, including reducing greenhouse gas emissions, lowering input energy costs, and enhancing energy security. Advances in science and technology have contributed to lowering the manufacturing costs of PV panels while increasing their efficiency. Additionally, modern control techniques have significantly improved the efficiency, flexibility, cost-effectiveness, and overall performance of solar power systems. Examples of such techniques include maximum power point tracking (MPPT), inverter control techniques, energy management and feedback control, and model predictive current control (MPCC) [28]. Figure 5 shows a control system of a solar power system.

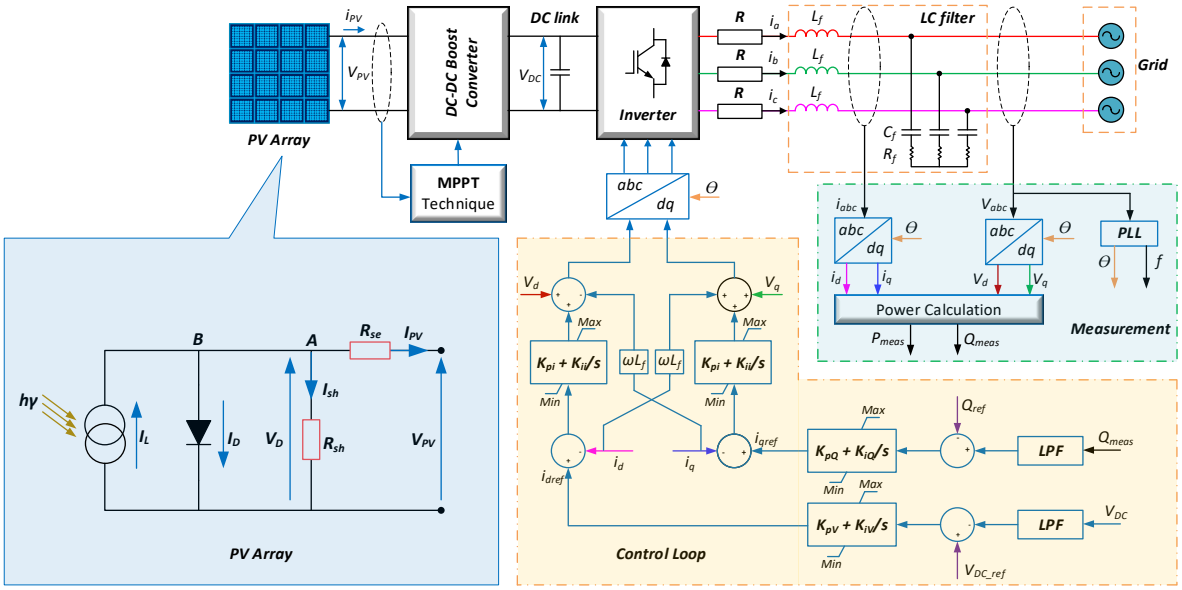


Figure 5. Control system of a solar power system.

According to Kirchhoff's current law, we have:

$$I_{PV} = I_L - I_D - I_{sh} \tag{4}$$

I_L is the photovoltaic current, independent of V_{PV} (or R_{se}), and is proportional to the incident light flux and the carrier diffusion length. It is linearly dependent on solar irradiance and also affected by temperature, as described by the following equation:

$$I_L = I_{SCr} + K_i\left(T - T_{ref}\right)\frac{G}{1000} \tag{5}$$

where I_{SCr} is the short-circuit current (under standard conditions), K_i is the short-circuit current temperature coefficient, T and T_{ref} are the operating temperature and reference temperature,

https://doi.org/10.31436/iiumej.v27i2.3683

respectively (in Kelvin), G is the solar irradiance on the panel surface (1000 W/m^2 is the nominal value) [28].

I_{sh} is the current flowing through the resistor R_{sh}, determined as follows:

$$I_{sh} = \frac{V_D}{R_{sh}} = \frac{V_{PV}+R_{se}I_{PV}}{R_{sh}} \quad (6)$$

I_D is the current flowing through the diode, determined as follows:

$$I_D = I_0\left(e^{\left(\frac{qV_D}{AKT}\right)} - 1\right) \quad (7)$$

where: q is the charge of an electron, $1.6.10^{-19}$C; K is the Boltzmann constant, $1.380649.10^{-23}$J/K; T is the temperature of the photovoltaic cell in Kelvin (K).

I_0 is the reverse saturation current of the diode, determined by:

$$I_0 = I_{rs}\left[\frac{T}{T_{ref}}\right]^3 \exp\left[\left(q\frac{E_{g0}}{Ak}\right)\left(\frac{1}{T_{ref}} - \frac{1}{T}\right)\right] \quad (8)$$

I_{rs} is the saturation current of the diode, determined by:

$$I_{rs} = \frac{I_{SCr}}{[\exp\left(\frac{qV_{0C}}{N_S kAT}\right)-1]} \quad (9)$$

Combining equations (4), (6), and (7), we obtain the characteristic equation as follows:

$$I_{PV} = I_L - I_0\left(e^{\left(\frac{qV_D}{AKT}\right)} - 1\right) - \frac{V_{PV}+R_{se}I_{PV}}{R_{sh}} \quad (10)$$

From the characteristic equation, we can deduce:

$$I_{PV} = \frac{I_L - I_0\left(e^{\left(\frac{qV_D}{AKT}\right)}-1\right) - V_{PV}/R_{sh}}{1+(R_{se}/R_{sh})} \quad (11)$$

where A is the ideality factor of the solar cell. In the ideal case: $R_{se} \approx 0$ (ideal series resistance) and $R_{se} \sim \infty$ (ideal shunt resistance).

2.3. Battery energy storage system modeling

A battery energy storage system (BESS), as shown in Figure 6, typically consists of three main components:

- **Battery Cell Arrays:** These consist of multiple battery cells connected in series to form rows, with these rows connected in parallel to assemble a complete cluster.
- **Power Conversion System (PCS):** This is a power electronic converter responsible for inverting DC power into AC power for loads or the grid, or rectifying AC power into DC power when charging the battery array [2, 15, 29]. The PCS operates based on control signals from the control and monitoring system.
- **Control and Monitoring System:** Serving as the core of the BESS, this system adjusts voltage, frequency, changes power output, and performs functions that stabilize the microgrid.

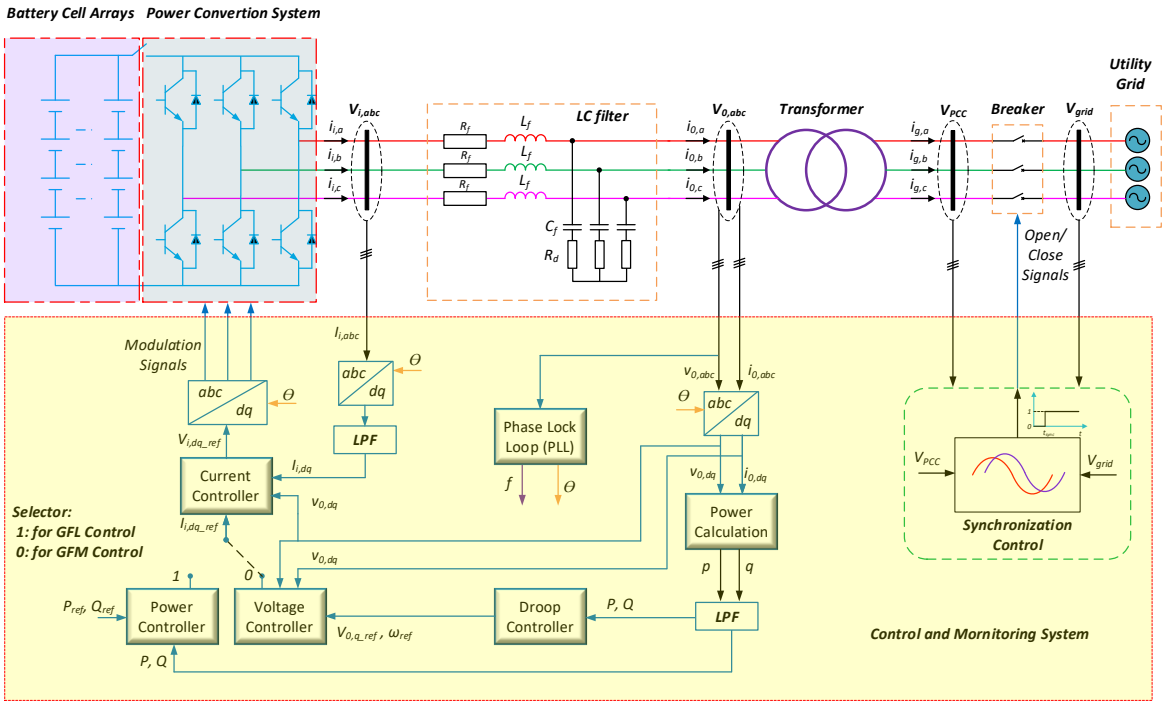


Figure 6. Control system of BESS.

The states of a battery energy storage system are defined as:

$$SoC_{min} \leq SoC(t) \leq SoC_{max} \tag{12}$$

where SoC_{max} is the fully charged state, equivalent to the rated storage capacity (C_n) and proportional to the total number of battery strings (N_b), the number of battery modules (N_{bs}), and the battery capacity (C_b).

$$C_n = C_b \left(\frac{N_b}{N_{bs}}\right) \tag{13}$$

SoC_{min} is the lowest allowable energy state of the battery energy storage system, calculated as:

$$SoC_{min} = (1 - DoD).SoC_{max} \tag{14}$$

where $DoD(\%)$ is the depth of discharge, i.e., the maximum allowable discharge percentage of the battery.

The no-load voltage of the battery (V_{nl}) is given by:

$$V_{nl} = V_0 - V_k \frac{1}{SoC} + A^{-BC_n(1-SoC)} \tag{15}$$

where V_{nl} is the no-load voltage of the battery (V), V_0 is the constant voltage of the battery (V), V_k is the polarization voltage of the battery (V), SoC is the state of charge of the battery (%), C_n is the capacity of the battery (Ah), and A, B are empirical constants.

Similar to the power converters used for distributed energy sources, the power converter of a BESS enables the conversion of active power (P) and reactive power (Q) across all four quadrants (*i.e.,* allowing independent absorption or generation of P and Q) within the rated current limits of the power electronics system.

The voltage and current values injected into the grid by the inverter are measured at the $V_{0,abc}$ node (located behind the LC filter). The three-phase voltage and current signals ($v_{0,abc}$

https://doi.org/10.31436/iiumej.v27i2.3683

and $i_{0,abc}$) collected are then passed through the Park Transform block (abc/dq) to convert them into voltage and current signals in the two-dimensional dq coordinate system ($v_{0,dq}$ and $i_{0,dq}$). The instantaneous power p, q injected into the system by the BESS are calculated in the power calculation block as:

$$p = \frac{3}{2}\left(v_{0,d}i_{0,d} + v_{0,q}i_{0,q}\right) \tag{16}$$

$$q = \frac{3}{2}\left(v_{0,d}i_{0,q} - v_{0,q}i_{0,d}\right) \tag{17}$$

To obtain the average values of active power P and reactive power Q, the instantaneous values p, q must be passed through a low-pass filter (LPF), in which $LPF \rightarrow \frac{\omega_c}{s+\omega_c}$.

$$\dot{P} = -P\omega_c + \frac{3}{2}\left(v_{0,d}i_{0,d} + v_{0,q}i_{0,q}\right) \tag{18}$$

$$\dot{Q} = -Q\omega_c + \frac{3}{2}\left(v_{0,d}i_{0,q} - v_{0,q}i_{0,d}\right) \tag{19}$$

In the Grid-Following (GFL) control mode, the active and reactive powers P, Q are sent to the power controller as shown in Figure 7. Here, the calculated P, Q values are compared with the reference values P_{ref}, Q_{ref}, and the errors are processed by a proportional-integral (PI) controller. The PI controller outputs the reference control signals i_{i,d_ref} and i_{i,q_ref} for the current controller.

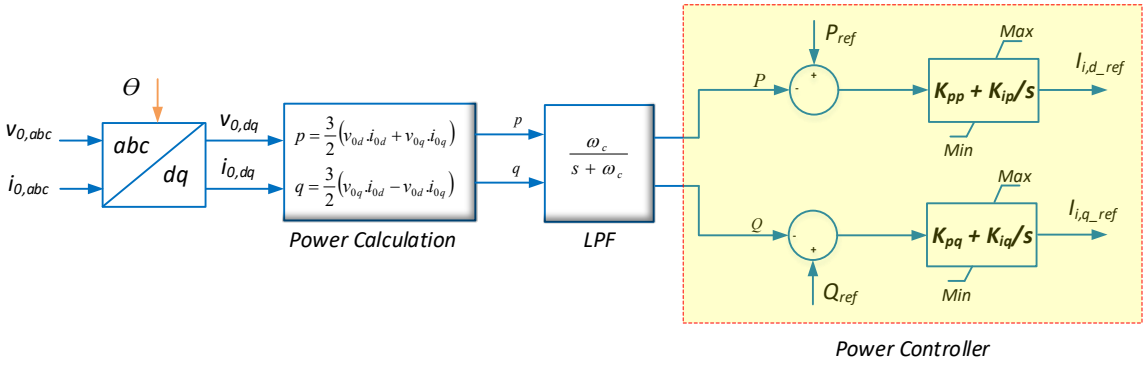


Figure 7. Power controller of BESS.

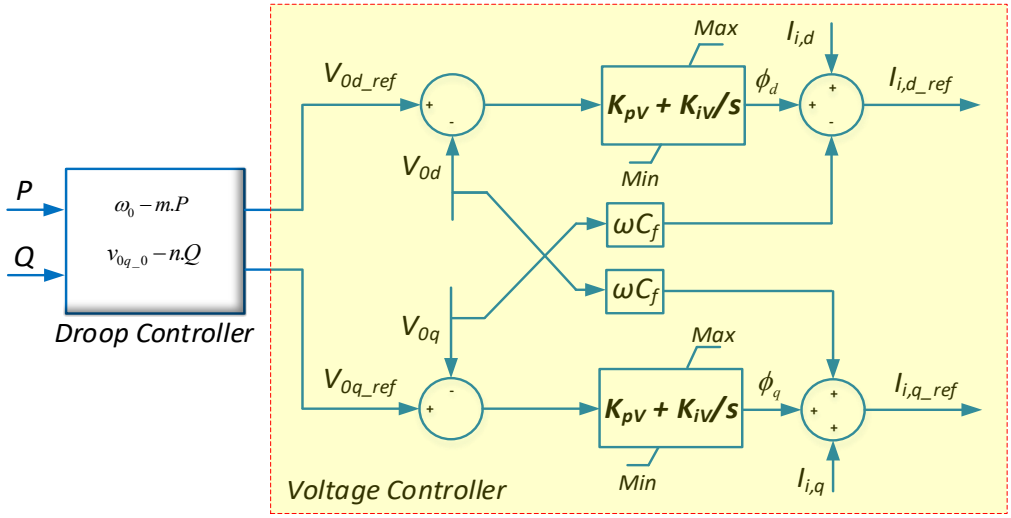


Figure 8. Voltage controller of BESS.

https://doi.org/10.31436/iiumej.v27i2.3683

In the Grid-Forming (GFM) control mode, the active and reactive powers P, Q are sent to the droop controller as shown in Figure 8. Here, P, Q are referenced to calculate and output the angular frequency ω_{ref} and the reference voltage V_{0,dq_ref}, which are provided to the voltage controller. The reference voltage V_{0,dq_ref} is then compared with the measured voltage $V_{0,dq}$ at node $V_{0,abc}$, and the error is processed by a proportional-integral (PI) controller. The PI controller outputs the reference control signals $i_{i,dref}$ and $i_{i,qref}$ for the current controller.

The droop controller equations are:

$$\omega_{ref} = \omega_n - m_p(P - P_0) \text{ or } \Delta\omega_{ref} = -m_p\Delta P \tag{20}$$

$$V_{0,d_ref} = V_{0,d_n} - n_q(Q - Q_0) \text{ or } \Delta V_{0,d_ref} = -n_q\Delta Q \tag{21}$$

The equations of the voltage controller are:

$$\frac{d\phi_d}{dt} = v_{0,d_ref} - v_{0,d} \tag{22}$$

$$\frac{d\phi_q}{dt} = v_{0,q_ref} - v_{0,q} \tag{23}$$

$$i_{i,d_ref} = i_{0,d} - \omega_n C_f v_{0,q} + K_{pv}\left(v_{0,d_{ref}} - v_{0,d}\right) + K_{iv}\phi_d \tag{24}$$

$$i_{i,q_ref} = i_{0,q} + \omega_n C_f v_{0,d} + K_{pv}\left(v_{0,q_ref} - v_{0,q}\right) + K_{iv}\phi_q \tag{25}$$

The inverter-side current signal (measured before the LC filter) is measured at the $V_{i,abc}$ node. This three-phase signal is transformed using the Park Transform (abc/dq) into a two-dimensional dq-axis current signal ($i_{i,dq}$), which is then passed through a low-pass filter (LPF) and sent to the current controller as shown in Figure 9. At the controller, the measured inverter current signal $i_{i,dq}$ is compared with the reference value i_{i,dq_ref} obtained from the previous control stage. The error between these values is processed by a proportional-integral (PI) controller. The reference output voltage signal v_{i,dq_ref} is determined by incorporating feedforward voltage compensation and cross-coupling compensation, where ωL_f represents the filter's equivalent reactance. This reference voltage signal is then passed through the inverse Park Transform (dq/abc) to generate the pulse modulation signals for the inverter.

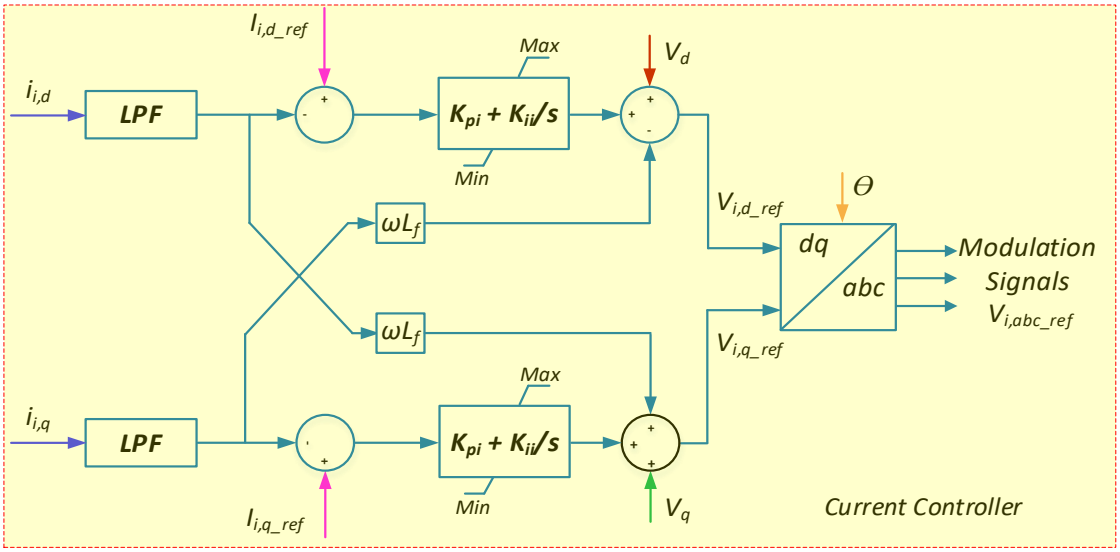


Figure 9. Current controller of BESS.

https://doi.org/10.31436/iiumej.v27i2.3683

$$\frac{d\gamma_d}{dt} = i_{i,d_ref} - i_{i,d} \quad (26)$$

$$\frac{d\gamma_q}{dt} = i_{i,q_ref} - i_{i,q} \quad (27)$$

$$v_{i,d_ref} = v_{0,d} - \omega_n L_f i_{i,q} + K_{pi}\left(i_{i,d_ref} - i_{i,d}\right) + K_{ii}\gamma_d \quad (28)$$

$$v_{i,q_ref} = v_{0,q} + \omega_n L_f i_{i,d} + K_{pi}\left(i_{i,q_ref} - i_{i,q}\right) + K_{ii}\gamma_q \quad (29)$$

The equations of the LC filters are:

$$\frac{di_{i,d}}{dt} = -\frac{R_f}{L_f} i_{i,d} + \omega i_{i,q} + \frac{1}{L_f} v_{i,d} - \frac{1}{L_f} v_{0,d} \quad (30$$

$$\frac{di_{i,q}}{dt} = -\frac{R_f}{L_f} i_{i,q} - \omega i_{i,d} + \frac{1}{L_f} v_{i,q} - \frac{1}{L_f} v_{0,q} \quad (31)$$

$$\frac{dv_{0,d}}{dt} = \omega v_{0,q} + \frac{1}{C_f} i_{i,d} - \frac{1}{C_f} i_{0,d} \quad (32)$$

$$\frac{dv_{0,q}}{dt} = -\omega v_{0,d} + \frac{1}{C_f} i_{i,q} - \frac{1}{C_f} i_{0,q} \quad (33)$$

2.4. Microgrid modeling

The microgrid is an integral component of the future smart grid, capable of operating autonomously in either islanded mode or grid-connected mode. In the event of a fault in the grid-connected mode, the microgrid can disconnect from the main grid and independently supply power to loads within the microgrid, as shown in Figure 10. The components of the microgrid studied in this paper include AC loads, distributed generation sources (including wind farms and solar power systems), and, notably, a battery energy storage system [11].

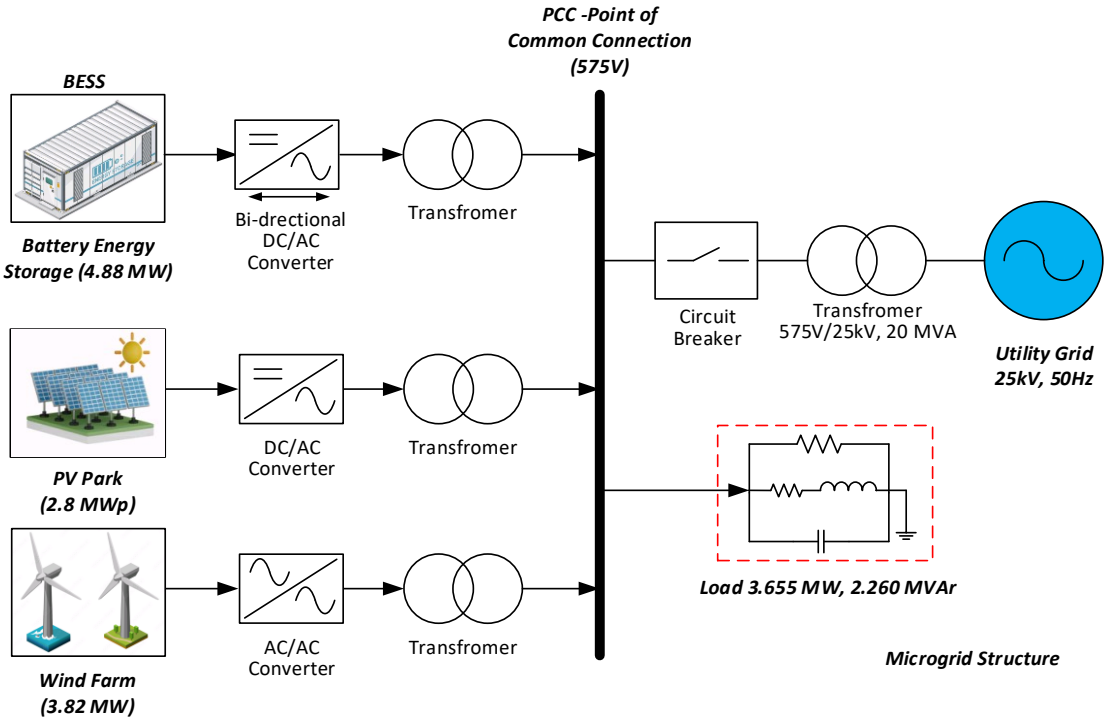


Figure 10. Microgrid structure.

The paper investigates and simulates the microgrid in two cases: grid-connected mode and islanded modes, corresponding to different scenarios, to evaluate the control performance of

inverters for both the grid-following and grid-forming functions of the BESS in stabilizing the microgrid. An inverter operating with a grid-following control function is treated as a controlled current source that injects current into the grid, as shown in Figure 11. The main components include an impedance connected in parallel with a current source. In grid-following mode, the inverter is synchronized with the grid voltage through a phase-locked loop (PLL). The active and reactive power injected into the grid is controlled by adjusting the reference current values I_{dq_ref}. In this mode, the inverter only controls the power output (active and reactive power) and does not participate in regulating the grid voltage or frequency.

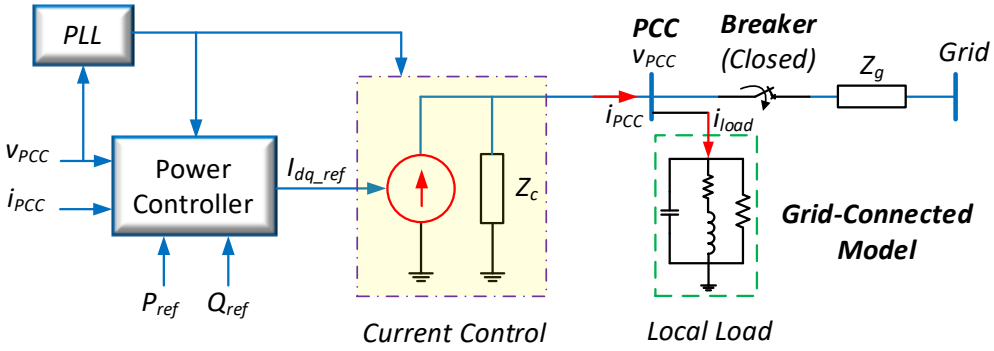


Figure 11. Grid-following control system.

In grid-connected mode, the microgrid's voltage and frequency are regulated by the synchronous generators of the main grid. However, in islanded mode, the microgrid must independently regulate its voltage and frequency to ensure stable operation, as shown in Figure 12. Unlike the grid-following control function, an inverter with grid-forming control is treated as a controlled voltage source connected in series with an output impedance, functioning similarly to a synchronous generator. In islanded mode, the inverter operates in grid-forming mode, where the voltage and frequency at the point of common coupling (PCC) are measured and compared against the reference values to control the inverter's output voltage.

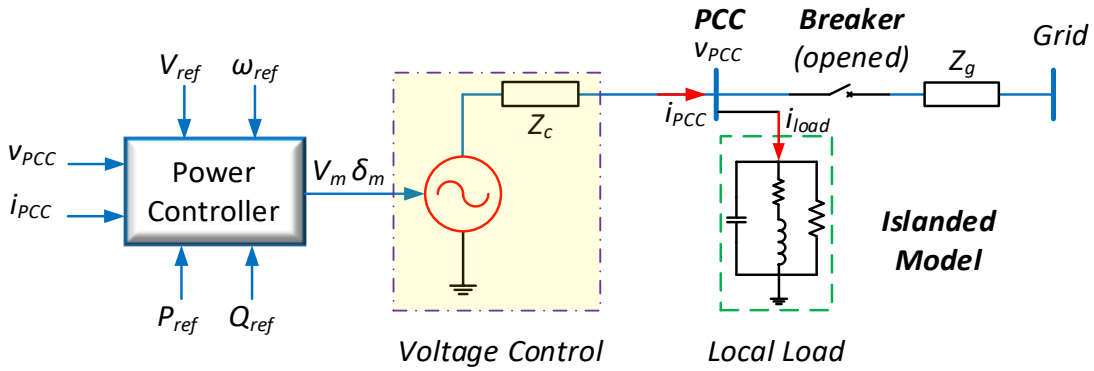


Figure 12. Grid-forming control system.

In the proposed microgrid model presented in the paper (including solar power systems, wind farms, and energy storage systems), the inverters used in the simulation include both grid-following (GFL) and grid-forming (GFM) types. Solar power systems and wind farms utilize grid-following inverters, while the energy storage system employs grid-forming inverters. This configuration accurately reflects the current state of modern power grids. Previously, the control of wind farms and solar power systems primarily focused on delivering only active power ($PF = 1$) due to the economic benefits for investors. However, the integration of BESS

into the grid has become essential in both present and future contexts to ensure reliable grid operation and stability, as well as to maximize the effective utilization of renewable energy sources.

2.5. Proposed Control Strategy

The proposed control strategy distinguishes itself by integrating both grid-following (GFL) and grid-forming (GFM) inverter control modes within a unified Battery Energy Storage System (BESS) framework. This hybrid scheme enables flexible mode switching based on the grid's operational state, supporting both grid-connected and islanded operations.

Unlike conventional approaches that often regulate voltage and frequency separately, the proposed strategy enables simultaneous control of both voltage and frequency within the simulation domain, improving the microgrid's dynamic behavior during transition events such as load changes and grid reconnection. These dynamics are captured through time-based simulations to evaluate the system's stability under varying operating conditions.

Another novel aspect lies in the integration of representative renewable energy fluctuations, including controlled PV curtailment and wind speed variations, into the simulated operational and fault scenarios. This interaction between fluctuating renewable inputs and control performance is rarely investigated in depth in previous simulation studies, as illustrated in the comparative analysis (Table 1).

Overall, this simulation-based control strategy positions the BESS not only as a stabilizing component but also as a flexible interface that enhances microgrid robustness and adaptability across diverse modeled scenarios.

3. Simulation Results and Discussion

3.1. Case Study 1

In Case Study 1, the microgrid operates in grid-connected mode and transitions to islanded mode. The operational characteristics of the components are as follows: the solar power source generates a fixed power of $2.8\,MW$, the wind power source generates a fixed power of $3.82\,MW$, the load consumes a fixed active power of $4.005\,MW$ and the reactive power of $2.31\,MVAr$, the BESS operates in grid-following mode (GFL) and transitions to grid-forming mode (GFM) before reconnecting to the grid. The simulation in Case Study 1 is conducted over a 10-second period in Simulink, using the specified parameters. The input data for the distributed generation sources (wind and solar) and load power are as specified above. At $t = 1\,s$, a transient fault occurs on the transmission line, and the microgrid begins operating in islanded mode. Once the transient fault is cleared, the microgrid is commanded to reconnect to the grid at $t = 5\,s$. The circuit breaker is reclosed at $t = 8\,s$.

From Figure 13, we can see the microgrid power graph for Case Study 1. At $t = 1\,s$, the microgrid begins to transition to islanded mode, causing transient power fluctuations as the system changes state. The BESS starts absorbing power (i.e., enters the charging state) to balance the grid's power, and the power flow through the PCC becomes zero from this point onward. The load power slightly decreases compared to the pre-islanding stage, as the microgrid no longer exports power to the main grid via the PCC, resulting in a voltage drop within the microgrid. The power output of other distributed generation sources remains nearly unchanged after approximately two seconds of oscillation. At $t = 5\,s$, when the command to reconnect to the grid is issued, the BESS also begins to adjust the voltage magnitude, frequency,

https://doi.org/10.31436/iiumej.v27i2.3683

and phase angle to synchronize with the main grid. Consequently, the BESS's reactive power output increases to compensate for the voltage deficit that occurred during islanded operation.

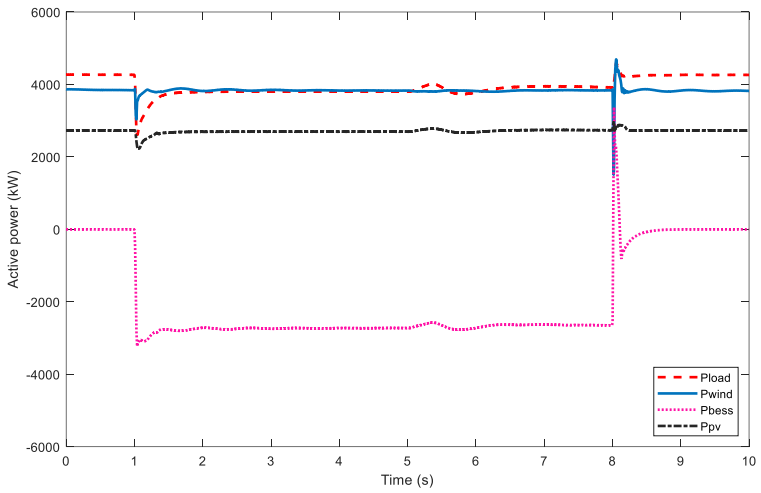


Figure 13. The active powers of the Case Study 1.

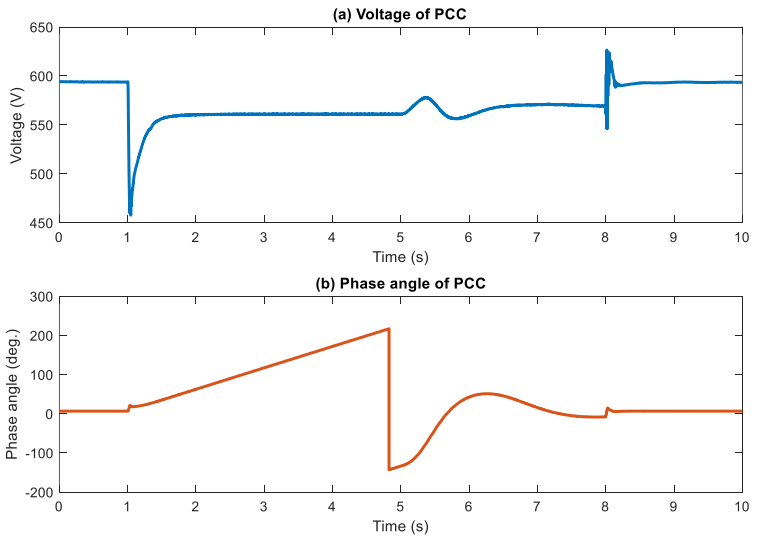


Figure 14. The voltage and phase angle at the PCC of Case Study 1.

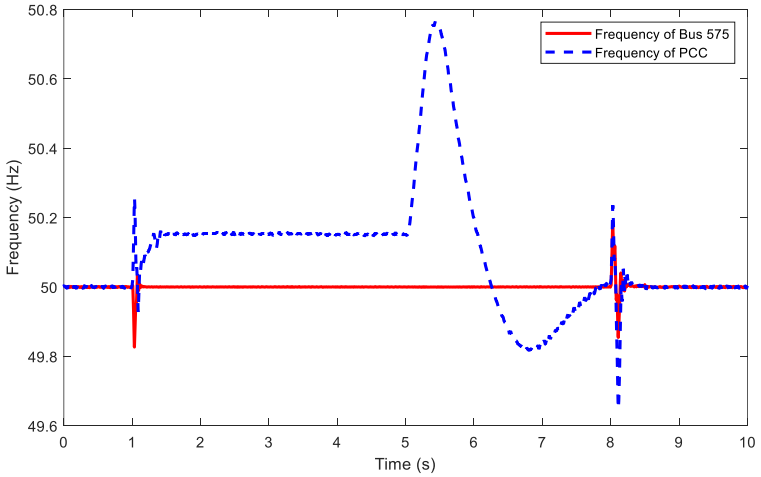


Figure 15. The frequency of the PCC and bus 575 in Case Study 1.

From Figures 14 and 15, it is easy to observe that the voltage at the PCC node before islanding is relatively high, approximately $594\,V$, because the power generated by the distributed sources is significantly greater than the load demand, and the excess power is exported to the main grid. Within the microgrid, the frequency remains approximately $50\,Hz$, essentially unchanged, as it is maintained in grid-connected mode. Upon islanding, the microgrid frequency increases slightly to around $50.15\,Hz$ due to the surplus generation, and the voltage, maintained by the BESS, drops to about $561\,V$ owing to the droop characteristics of the BESS. At $t = 5\,s$, when the reconnection command is issued, the frequency, phase angle, and voltage begin to adjust to match the grid values ($575\,V$ and $50\,Hz$), and synchronization with the grid is completed at $t = 8\,s$.

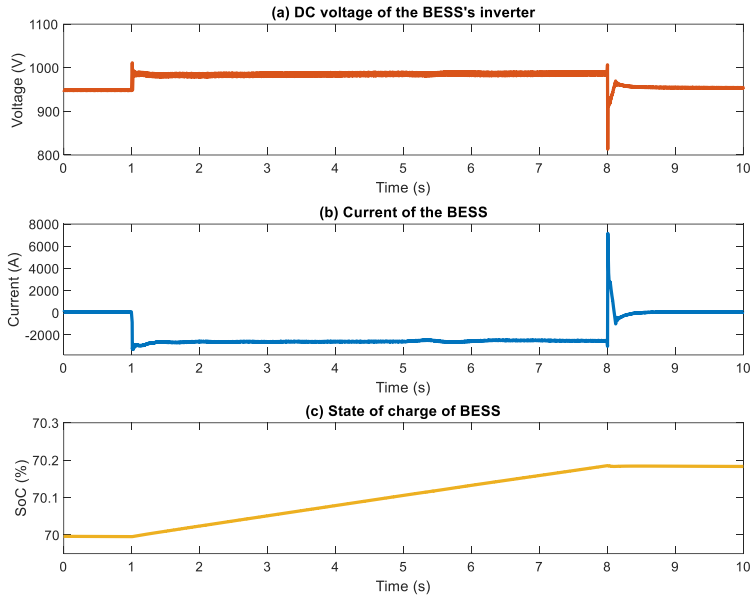


Figure 16. The BESS parameters of Case Study 1.

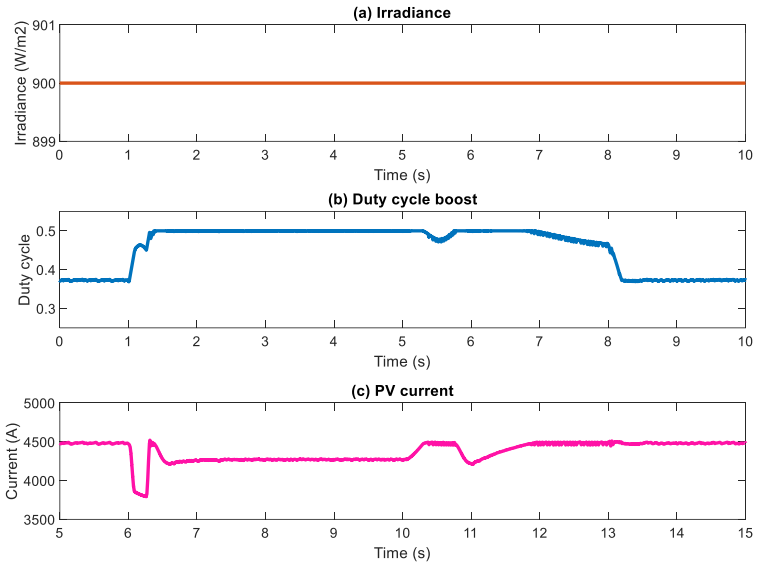


Figure 17. The solar power system parameters of Case Study 1.

The DC voltage, battery current, and state of charge (SoC) of the BESS are illustrated in Figure 16. Before the islanding event, the BESS neither absorbs nor delivers power, so its DC voltage is maintained at the nominal level (approximately 950 *V*). However, during islanding,

the BESS absorbs active power, causing its DC voltage to increase slightly depending on the charging dynamics and battery characteristics (approximately 985 *V*). The battery current becomes negative during islanding because the BESS operates in charging mode, increasing its SoC during this period.

In Case Study 1, the power output of the solar and wind sources remains constant. However, the parameters of these two renewable energy sources are also affected and exhibit noticeable transient fluctuations (as detailed in Figures 16-19). In this context, the role of the BESS, specifically the grid-forming inverter, has proven effective in controlling the microgrid's voltage and frequency. This ensures the microgrid can operate independently during fault conditions and facilitates reconnection to the main grid despite fluctuations caused by integrating renewable energy sources in a weak-grid environment.

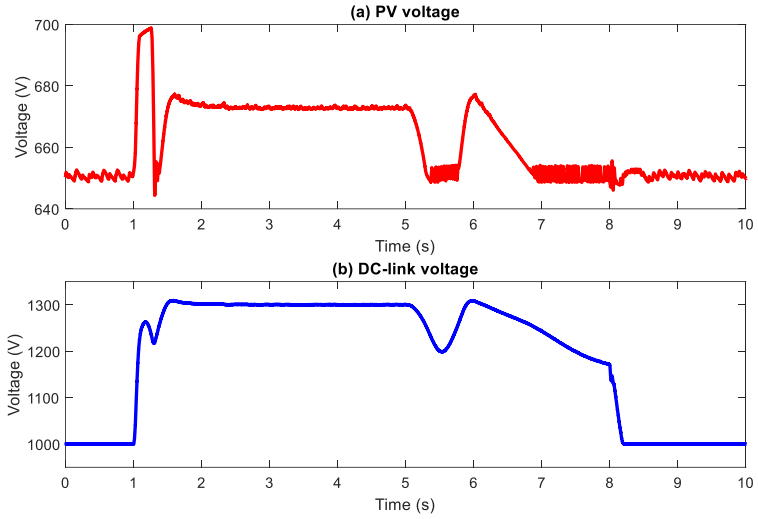


Figure 18. The solar power system parameters of Case Study 1.

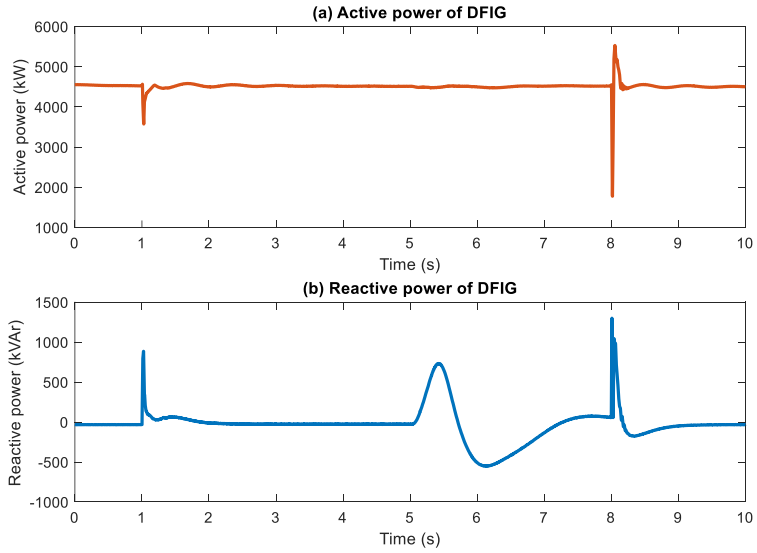


Figure 19. The active and reactive powers of the DFIG of Case Study 1.

In this case, the BESS, in addition to demonstrating the ability to independently and stably control the grid in islanded mode, also demonstrates enhanced dynamic performance during reconnection to the main grid while maintaining seamless grid operation. From Figures 14 and 15, it can be observed that during islanded mode, the PCC voltage is lower than the grid voltage

(561, cap V), and the frequency is slightly higher than the nominal grid frequency (50.15, cap Hz). However, when reconnection to the main grid is required, the BESS gradually adjusts these parameters toward the nominal grid values (575 V, 50 Hz) by actively regulating PQ exchange, as illustrated in Figure 14. This control capability is achieved despite the continuous fluctuations in the output parameters of the renewable energy sources during the grid synchronization phase (as shown in Figures 17-20).

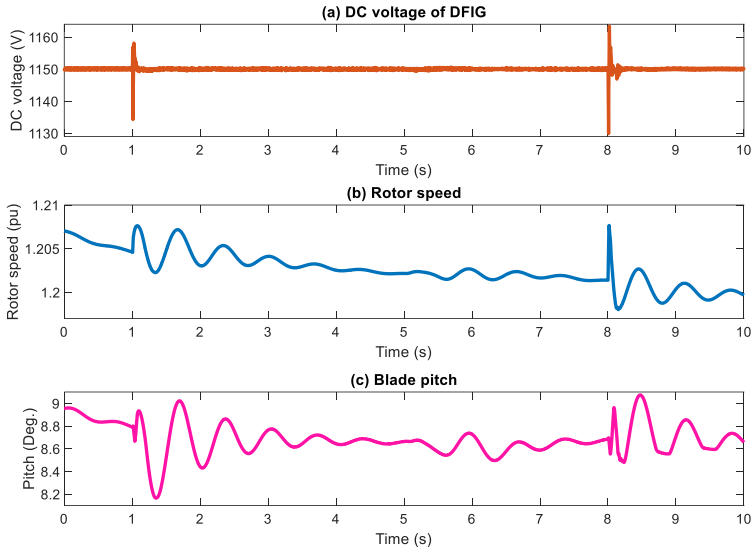


Figure 20. The DFIG parameters of Case Study 1.

3.2. Case Study 2

In Case Study 2, the microgrid operates in islanded mode, with the operating characteristics of the components as follows: the solar power source is controlled to vary its output power, the wind power source output fluctuates due to changes in wind speed, and the load consumes a constant power of 4.005 MW and 2.31 $MVAr$, the BESS operates in grid-forming mode (GFM). The input data for the output power of distributed generation sources (wind and solar) and the load power are as mentioned above. At $t = 1\ s$, the circuit breaker trips, and the microgrid transitions to islanded mode. During the islanded operation, the power outputs of distributed generation sources will fluctuate. For the solar power source, the initial output power is 2.8 MW; at $t = 3\ s$, it is reduced to 1 MW; at $t = 5\ s$, it drops to zero (0 MW), and at $t = 7\ s$, it increases again to 2 MW. For the wind power source, the initial wind speed is 15 m/s; at $t = 2\ s$, the wind speed decreases to 6 m/s, and at $t = 7\ s$, it rises to 12 m/s. To fully evaluate the functionality of BESS and the grid-forming inverter, Case Study 2 extends the analysis beyond Case Study 1 by examining additional operating scenarios of the grid-forming inverter. While the output power of renewable energy sources fluctuates in grid-connected mode, the BESS operates only in grid-following mode. In this case, the BESS assumes the grid-forming role within the microgrid to ensure stable operation. As shown in Figure 21, the microgrid power graph for Case Study 2 is illustrated. During the simulation of Case Study 2, after transitioning to islanded mode, the BESS participates in both absorbing and injecting active and reactive power to maintain the microgrid's frequency and voltage within permissible limits.

The steady-state voltage at the PCC node (Figure 22) during the islanded operation phase fluctuates between 562 and 567 V. The voltage ripples in this phase are caused by the power control of renewable energy sources (solar power). While variations in the output power of the solar source result in significant voltage oscillations, changes in wind power output are

https://doi.org/10.31436/iiumej.v27i2.3683

smoother and less abrupt. This difference arises from the varying sensitivity and inherent inertia of each energy source. At $t = 7\ s$, when the output power of renewable energy sources reaches its lowest point, the transient voltage of the microgrid dips to a minimum of $550\ V$.

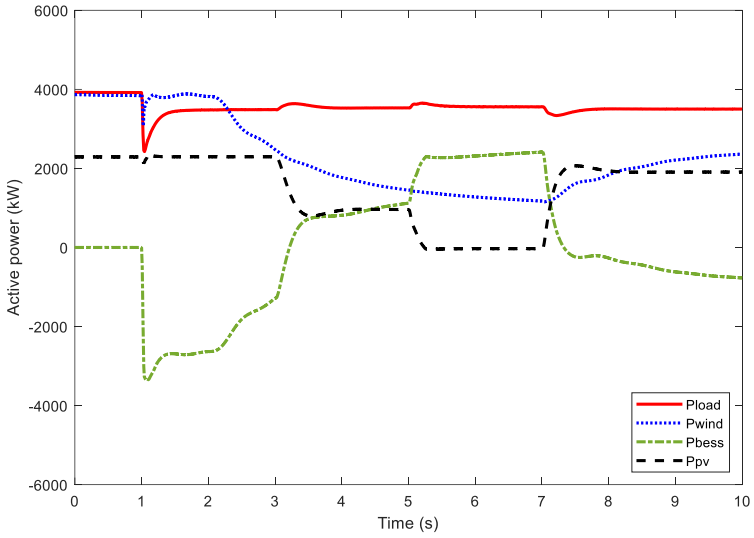


Figure 21. The active powers of Case Study 2.

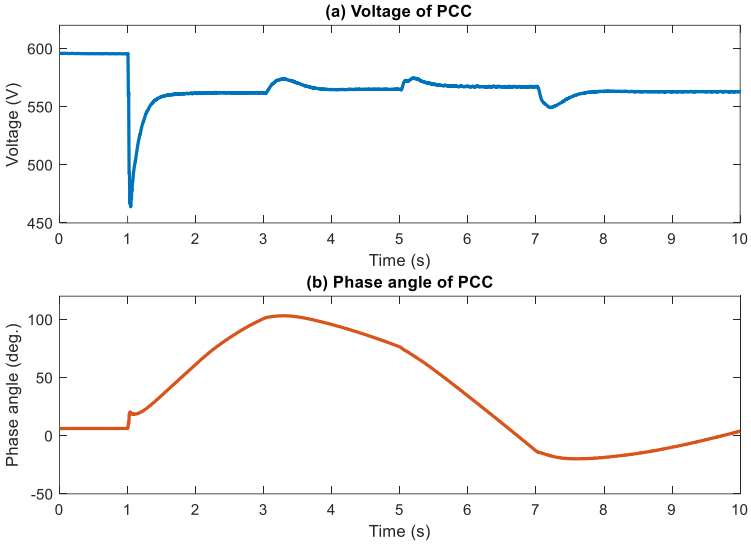


Figure 22. The voltage and phase angle at the PCC of Case Study 2.

In Figure 23, the grid frequency fluctuates within the range of $49.85 - 50.15\ Hz$ during the islanded operation phase. The frequency values accurately reflect the power output of renewable energy sources. Figure 24 illustrates the detailed parameters of the BESS under grid power fluctuations caused by renewable energy sources. The state of charge (SoC) of the battery clearly shows the alternating charging and discharging states of the BESS, which help stabilize the grid. Figures 25 and 26 depict the variations in solar power parameters under Case Study 2. In this case, the radiation value remains constant, and the solar power output is not controlled by maximum power point tracking (MPPT). Instead, it is forcibly curtailed or increased by adjusting the power control mode of the solar power control system.

https://doi.org/10.31436/iiumej.v27i2.3683

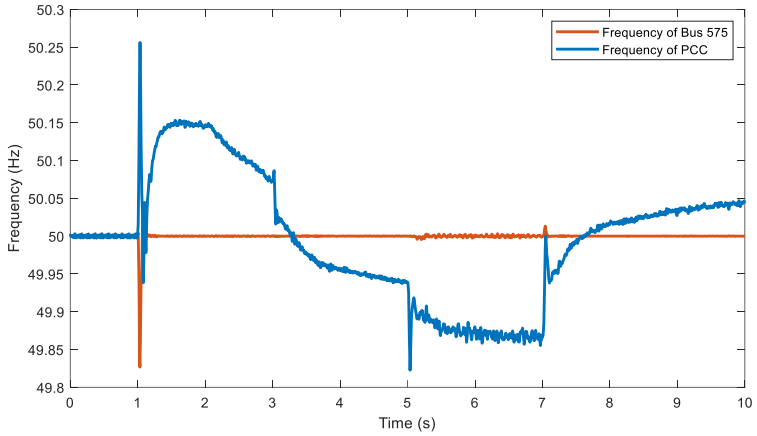


Figure 23. The frequency of the PCC and bus 575 in Case Study 2.

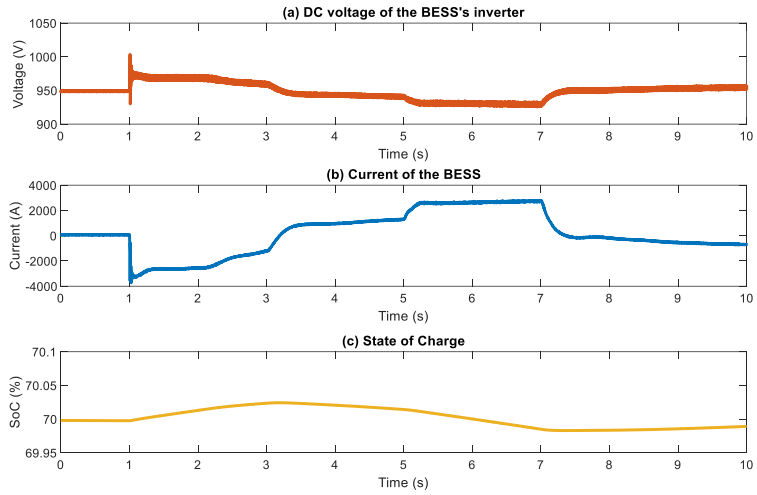


Figure 24. The BESS parameters of the Case Study 2.

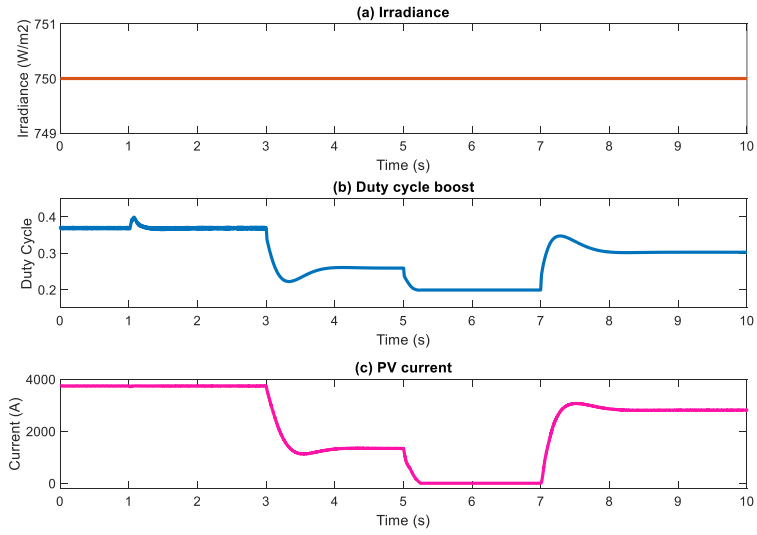


Figure 25. The solar power system parameters of Case Study 2.

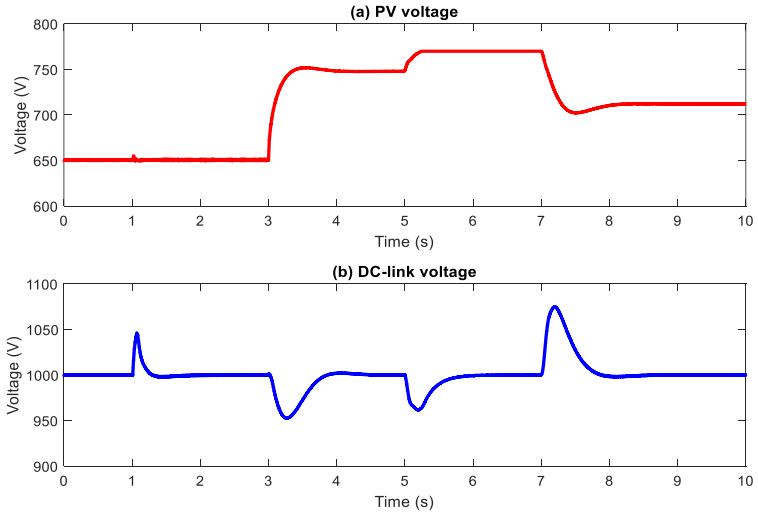


Figure 26. The solar power system parameters of the Case Study 2 (cont.)

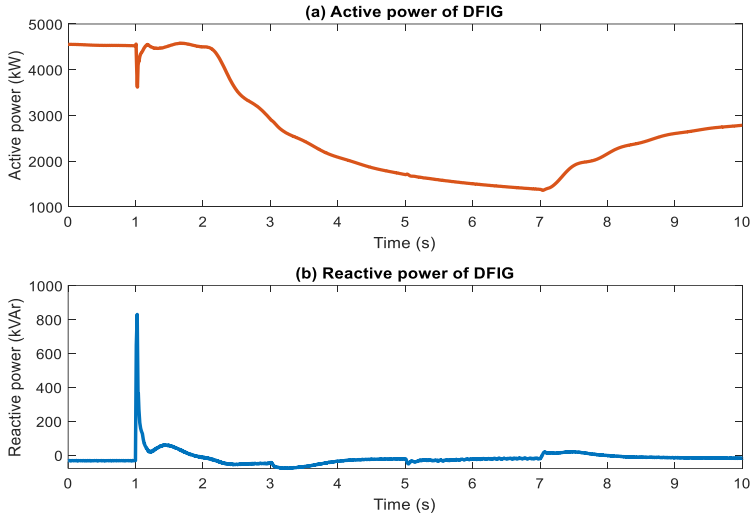


Figure 27. The active and reactive powers of the DFIG of Case Study 2.

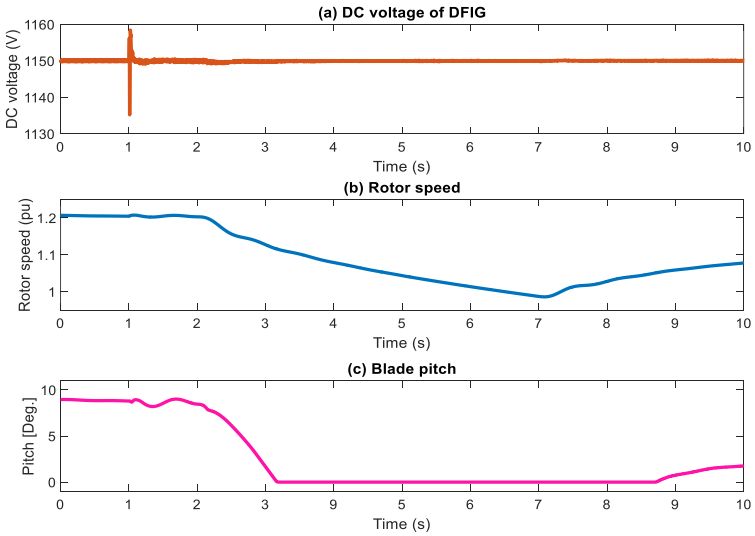


Figure 28. The DFIG parameters of Case Study 2.

https://doi.org/10.31436/iiumej.v27i2.3683

Figure 28 illustrates the self-regulation behavior of power output when the wind speed drops below the rated speed. The responses of wind turbines shift to sub-synchronous operation: the wind turbine speed decreases to the rated speed (1.0 pu), and the pitch angle is controlled to zero degrees to maximize power extraction under these operating conditions. For Case Study 2, the grid operates under varying power conditions of renewable energy sources; however, in this scenario, the grid operates in islanded mode. This case demonstrates the BESS's and the grid-forming inverter's capability to stabilize the grid during islanded operation. The voltage and frequency parameters accurately reflect the performance characteristics of the droop control method.

3.3. Case Study 3

Case Study 3 involves the microgrid transitioning from grid-connected operation to islanded mode. The components exhibit the following characteristics: the solar power source produces a fixed output of 2.8 MW, the wind power generates a fixed output of 3.82 MW, the load consumes a fixed power of 4.005 MW and 2.31 $MVAr$and the BESS operates in grid-following (GFL) mode and transitions to grid-forming (GFM) mode, eventually reconnecting to the grid. Input data for distributed generation (solar and wind) and load power are as described above. At $t = 1\ s$, a fault occurs on the transmission line, causing the microgrid to switch to the islanded mode. During islanded operation, at $t = 5\ s$, a three-phase-to-ground fault transient occurs at the PCC and lasts for 8 cycles (160 ms).

Figure 29 illustrates the power dynamics in the microgrid under Case Study 3. At $t = 1\ s$, the microgrid begins islanded operation, causing power disturbances due to the state transition. The BESS starts absorbing power (i.e., enters charging mode) to balance the microgrid's power, and the power flow through the PCC becomes zero. At $t = 5\ s$, when the short circuit occurs, all components in the microgrid experience significant disturbances in power, voltage, and frequency, as detailed in Figures 30 and 31. From Figures 30 and 31, it is evident that the voltage at the PCC drops to zero during the short circuit, and the microgrid frequency undergoes large fluctuations around 50 Hz. With the BESS intervention, the microgrid voltage quickly recovers to normal and stabilizes by approximately $t = 5.4\ s$. However, wind power remains oscillatory due to wind turbine inertia, causing the microgrid frequency to continue to fluctuate. During this period, to maintain system stability, BESS adjusts its operation to compensate for variations in wind power, eventually achieving stability at $t = 9\ s$.

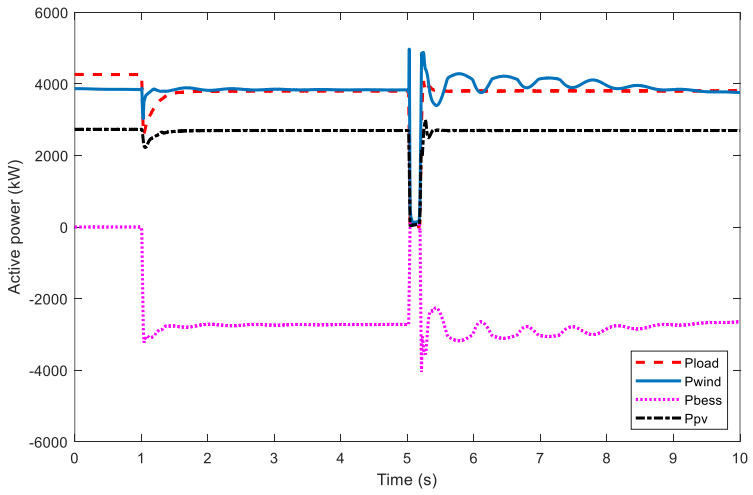


Figure 29. The active powers of the Case Study 3.

https://doi.org/10.31436/iiumej.v27i2.3683

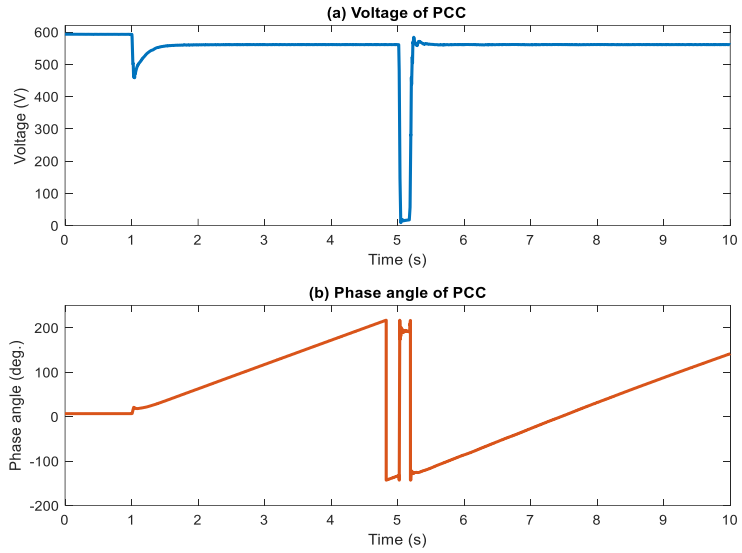


Figure 30. The voltage and phase angle at the PCC of Case Study 3.

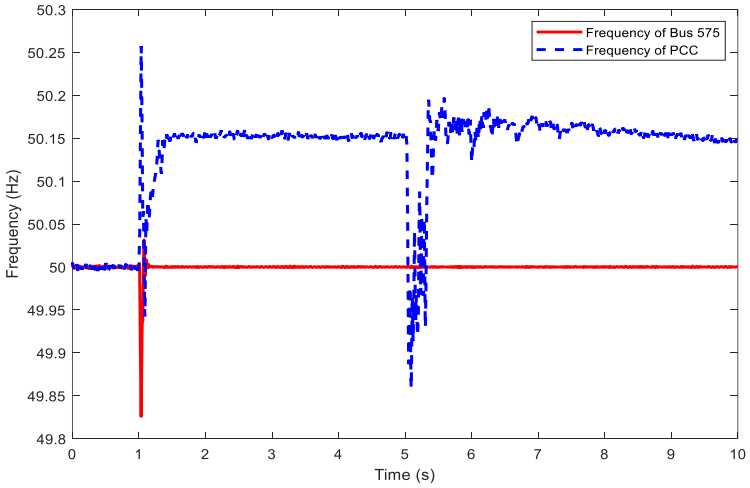


Figure 31. The frequency of the PCC and bus 575 in Case Study 3.

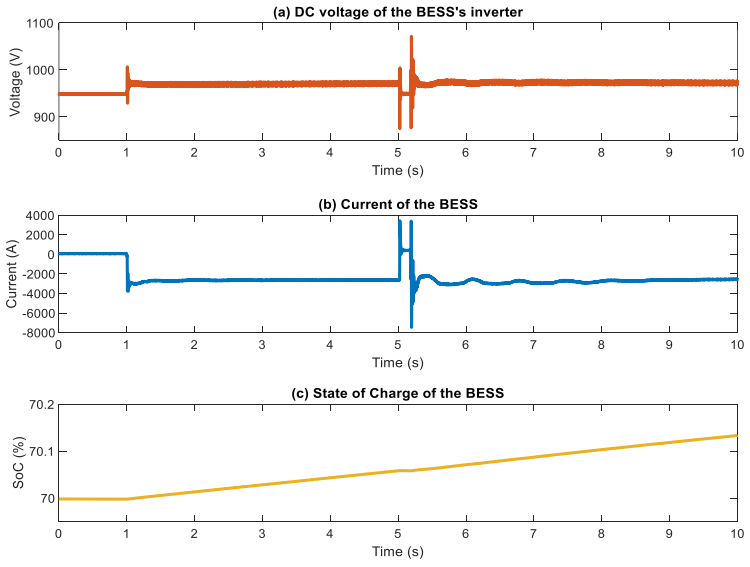


Figure 32. The BESS parameters of Case Study 3.

https://doi.org/10.31436/iiumej.v27i2.3683

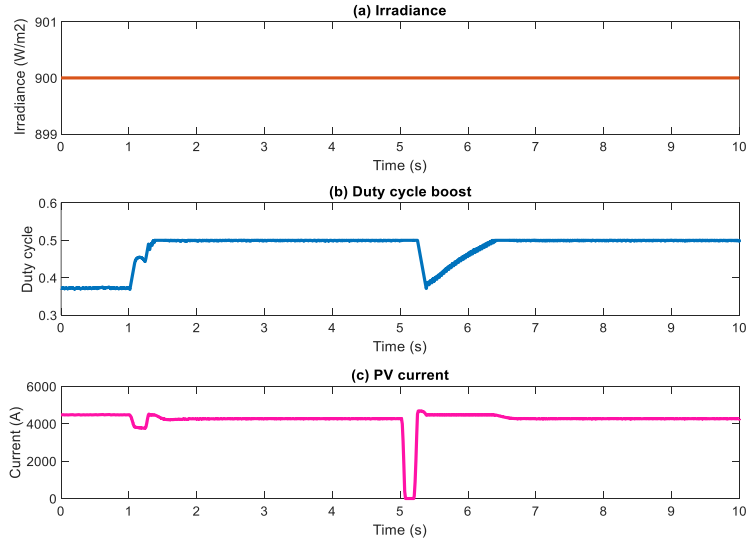


Figure 33. The solar power system parameters of Case Study 3.

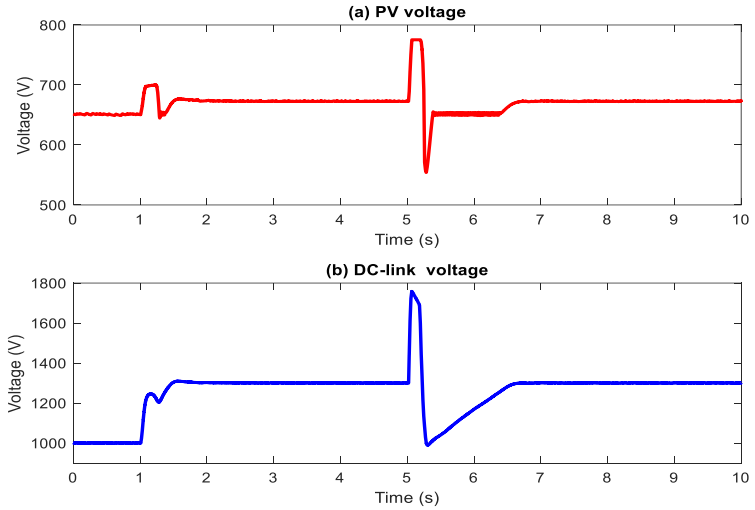


Figure 34. The solar power system parameters of the Case Study 3 (cont.)

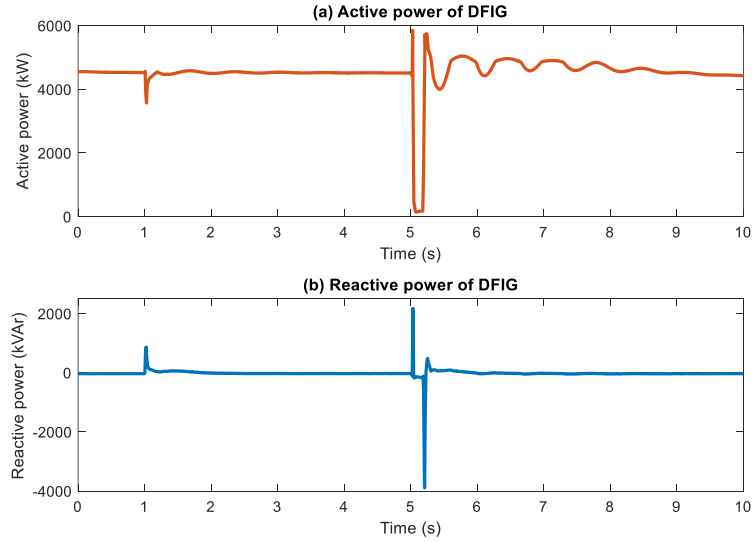


Figure 35. The active and reactive powers of the DFIG of Case Study 3.

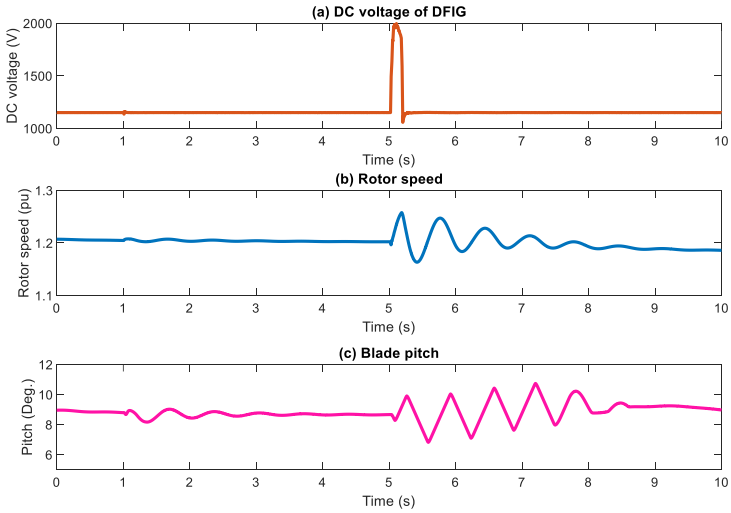


Figure 36. The DFIG parameters of Case Study 3.

From Figure 33**,** it is evident that the BESS parameters are disrupted at $t = 5\ s$ due to the fault event. However, the BESS rapidly restores stability and begins fine-tuning its control strategy after $t = 5.4\ s$. This behavior highlights the system's resilience and demonstrates the effectiveness of BESS in stabilizing the microgrid both during and after disturbances. Observing Figure 33, it can be seen that, during a fault, the photovoltaic panel current drops to zero, while the panel voltage and DC-Link voltage increase significantly. As a result, the output current of the solar power plant remains nearly unchanged, as illustrated in Figure 34 when a short circuit occurs at the Point of Common Coupling (PCC). While the parameters of solar power are only slightly affected by the short circuit at the Point of Common Coupling (PCC), the wind power output is significantly impacted, with the transient period lasting nearly 4 seconds before stabilizing, as shown in Figures 35 and 36. The prolonged transient behavior of wind power can be attributed to the wind turbine's inherent inertia and the limited response speed of the blade-pitch control system.

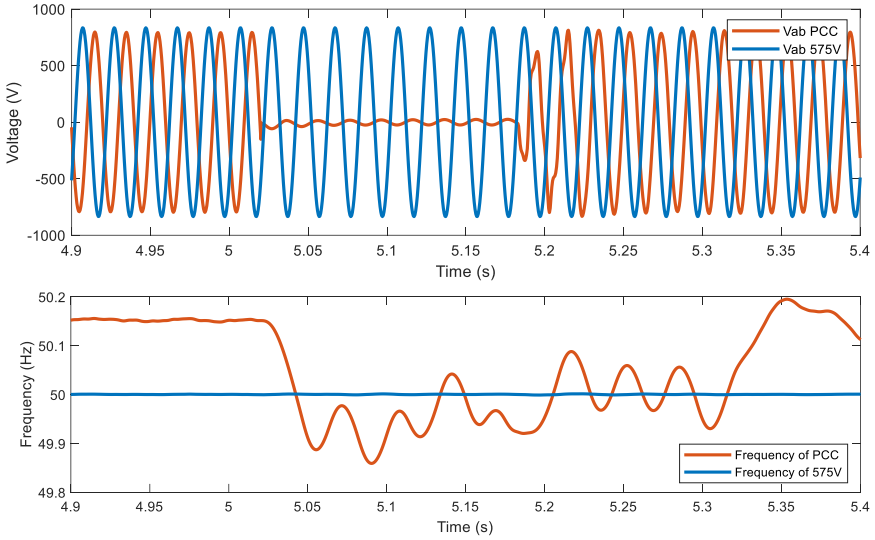


Figure 37. Voltage and frequency during a single-phase-to-ground fault.

https://doi.org/10.31436/iiumej.v27i2.3683

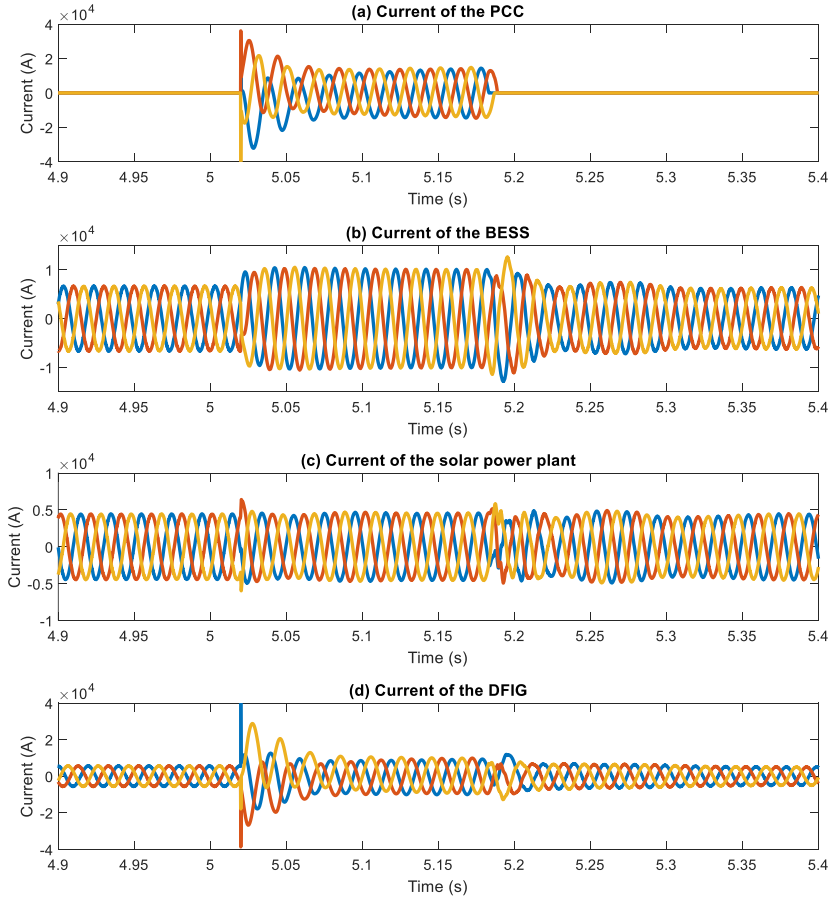


Figure 38. The current during a single-phase-to-ground fault.

Case Study 3 is an extension of Case Study 1, where the microgrid operates in islanded mode and simulates a short circuit lasting 160 ms to evaluate the responsiveness of the components within the grid. The simulation also serves as a basis for calculating and configuring relay protection settings for specific cases. Based on the responses of voltage, current, and frequency, as shown in Figures 37, 38, and 39, the Fault Ride-Through (FRT) capability of the elements in the microgrid is assessed.

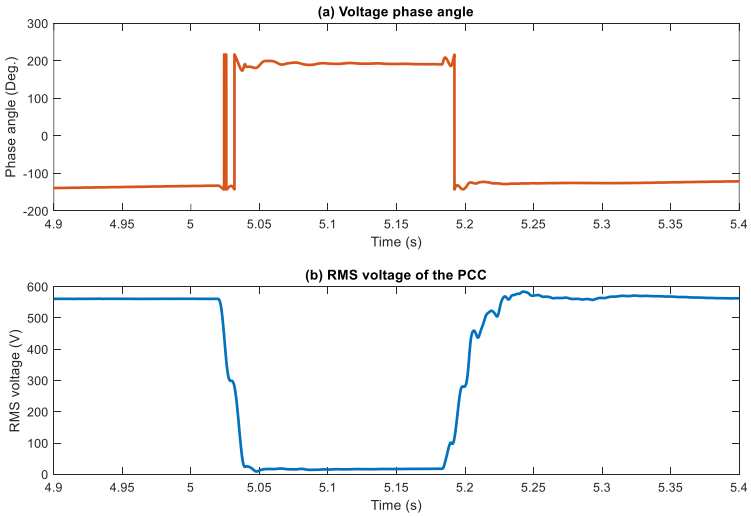


Figure 39. The voltage and phase angle at the PCC during a single-phase-to-ground fault.

https://doi.org/10.31436/iiumej.v27i2.3683

Table 2. Comprehensive performance summary of case studies

Parameter	Case Study 1	Case Study 2	Case Study 3
Objective	Evaluate islanding and reconnection under GFL↔GFM transitions	Assess stability under fluctuating RES in islanded mode	Evaluate fault ride-through capability during severe disturbances
Simulation Duration	10 s	10 s	10 s
Operation Mode	Grid-connected → Islanded → Reconnected	Grid-connected → Islanded (off-grid)	Grid-connected → Islanded with faults
Type of Disturbance	Manual islanding and reconnection	RES power fluctuation (PV, wind)	Three-phase-to-ground fault
BESS Mode	GFL → GFM → GFL	GFL → GFM	GFL → GFM
Main Control Strategy	Droop + synchronization	Droop + dynamic charge/discharge	Droop + FRT voltage restoration
Solar Power	2.8 MW (constant)	2.8 → 1 MW (t = 3s) → 0 MW (t = 5s) → 2 MW (t = 7s)	2.8 MW (constant)
Wind Power	3.82 MW (constant)	15 → 6 m/s (t = 2s) → 12 m/s (t = 7s)	3.82 MW (constant)
Load Demand	4.005 MW + 2.31 MVAr	4.005 MW + 2.31 MVAr	4.005 MW + 2.31 MVAr
Voltage at PCC	561 V (steady-state) and 556-578 V (during resynchronization)	562–567 V (steady state) and 550–575 V for voltage ripple during PV power changes.	0 V for an 8-cycle fault from t = 5 s, restored at t ≈ 5.4 s
Frequency at PCC	50.15 Hz (steady-state) 49.82-50.75 Hz (during resynchronization)	49.85 – 50.15 Hz	49.85 – 50.15 Hz
Voltage Recovery Time	3 s (during resynchronization)	Approximately 1 s for ripple (after each PV power drop)	Approximately 0.2s (after fault cleared)
Frequency Recovery Time	3 s (during resynchronization)	Due to RES power variations (wind and PV), the frequency fluctuates slightly; following a sudden PV power drop, recovery occurs within approximately 0.1–0.2 s.	Approximately 3.5–4 s (due to wind power inertia)
Reactive Power Behavior (Q)	Q injection for islanding	Dynamic Q fluctuation with P changes due to RES curtailments	Q distorted during fault, then stabilizes
SoC Behavior of BESS	Increase during islanded charging	Charging/discharging dynamically	interrupts during fault, then recharges
Stability Behavior	Smooth transition, damped oscillations	Stable, fast damping after disturbances	Slow recovery, affected by RES inertia
Synchronization Behavior	Successful and seamless	N/A	N/A
Grid Reconnection Time	From t = 5 s to t = 8 s	N/A	N/A
FRT Capability Shown	N/A	N/A	Full recovery from 0 V fault
BESS Action	Absorbs surplus power (charging) and supports Q	Maintains voltage and frequency dynamically	Provides voltage and frequency recovery
Impact of BESS	Ensures reconnection and stability	Maintains operation under severe RES changes	Enables recovery from major faults
Key Observation	Smooth GFL↔GFM transitions with reconnection	Effective GFM control under dynamic RES	Supports FRT and recovers grid stability

The comprehensive performance summary of case studies is shown in Table 2. Based on the comparative results, the BESS plays a vital role in enhancing grid stability across different operating conditions. In Case Study 1, the system achieves smooth GFL↔GFM transitions with successful reconnection, although the recovery time for voltage and frequency is relatively long. In Case Study 2, the BESS demonstrates strong dynamic performance under fluctuating RES conditions, with rapid voltage and frequency recovery within 0.1–0.2 s following sudden PV drops. This indicates the effectiveness of the droop control combined with dynamic charge/discharge. In Case Study 3, the BESS enables fault ride-through capability, restoring voltage in about 0.2 s, though frequency recovery is slower due to wind inertia. Each case highlights a different strength: reconnection support, dynamic stability, and fault tolerance. Overall, the results confirm that GFM-based BESS control significantly improves system resilience and operational flexibility.

4. CONCLUSIONS

Based on the simulation results, this paper presents a comprehensive overview of power grid operation with integrated renewable energy sources and highlights the critical role, benefits, and control strategies of Battery Energy Storage Systems (BESS) in microgrids. The study provides a detailed comparison between grid-connected and islanded microgrid operation modes, investigates both grid-following and grid-forming inverter control strategies, and analyzes the impacts of disturbances and fluctuations in distributed energy resources (DERs), particularly solar and wind generation. The findings demonstrate the effectiveness of BESS in stabilizing key grid parameters and maintaining system operation within permissible limits, while also emphasizing the potential of inverter technologies to enhance the controllability, flexibility, and resilience of microgrids. In this context, the study supports the transition toward highly inverter-based, and ultimately fully inverter-based, power grids that can maximize the utilization of clean and sustainable energy resources. In addition, the developed simulation model provides a practical framework for studying, developing, and designing new microgrids, testing operational parameters and grid stability, and evaluating the integration of renewable energy sources into existing power systems under established operational scenarios. Overall, this simulation framework lays the groundwork for future research on advanced control strategies and real-time hardware-in-the-loop testing, thereby supporting the further development and deployment of inverter-based microgrids in practical power system environments and broader grid modernization initiatives.

REFERENCES

[1] Masenge I, Mwasilu F. (2020) Hybrid solar PV-wind generation system coordination control and optimization of battery energy storage system for rural electrification. In Proceedings of the 2020 IEEE PES/IAS PowerAfrica, 25-28 August 2020, Nairobi, Kenya

[2] Zeng J, Zhang B, Mao C, Wang Y. (2006) Use of battery energy storage system to improve the power quality and stability of wind farms. Proceedings of the 2006 International Conference on Power System Technology, 22-26 October 2006, Chongqing, China.

[3] Energy Transition Investment Trends 2025 [https://about.bnef.com]

[4] Renewables 2025 Global Status Report 2025 [https://www.ren21.net/gsr-2025]

[5] Grid-forming Inverter Market Size, Share & Trends Analysis Report 2023-2028 [https://www.skyquestt.com/report/grid-forming-inverter-market]

[6] U.S. Energy Storage Monitor: Q4 2024 [https://cleanpower.org/news/u-s-energy-storage-monitor-q4-2024]

[7] De Siqueira LMS, Peng W. (2021) Control strategy to smooth wind power output using battery energy storage system: A review. Journal of Energy Storage, 35:102252.

[8] Gao F, Iravani MR. (2008) A control strategy for a distributed generation unit in grid-connected and autonomous modes of operation. IEEE Transactions on Power Delivery, 23(2):850-9.

[9] Babu VV, Roselyn JP, Nithya C, Sundaravadivel P. (2024) Development of Grid-Forming and Grid-Following Inverter Control in Microgrid Network Ensuring Grid Stability and Frequency Response. Electronics, 13(10):1958.

[10] Rathnayake DB, Akrami M, Phurailatpam C, Me SP, Hadavi S, Jayasinghe G. (2021) Grid forming inverter modeling, control, and applications. IEEE Access, 9:114781-807.

[11] Uddin M, Mo H, Dong D, Elsawah S, Zhu J, Guerrero J. (2023) Microgrids: A review, outstanding issues and future trends. Energy Strategy Reviews, 49:101127.

[12] Roy P, He J, Zhao T, Singh Y. (2022) Recent advances of wind-solar hybrid renewable energy systems for power generation: a review. IEEE Open Journal of the Industrial Electronics Society, 3:81-104.

[13] Kafle L, Ni Z, Tonkoski R, Qiao Q. (2016) Frequency control of isolated micro-grid using a droop control approach. Proceedings of the 2016 IEEE International Conference on Electro Information Technology (EIT), Grand Forks, 19-21 May 2016, Grand Forks, ND, USA.

[14] Poonahela I, Krama A, Bayhan S, Fesli U, Shadmand MB, Abu-Rub H. (2023) Hierarchical model-predictive droop control for voltage and frequency restoration in AC microgrids. IEEE Open Journal of the Industrial Electronics Society, 4:85-97.

[15] Zulkifly Z, Yusoff SH, Tumeran NL, Razali NSI. (2023) Battery Energy Storage System (BESS) Modeling for Microgrid. IIUM Engineering Journal, 24(1):57-74.

[16] Li X, Hui D, Lai X. (2013) Battery energy storage station (BESS)-based smoothing control of photovoltaic (PV) and wind power generation fluctuations. IEEE Transactions on Sustainable Energy, 4(2):464-73.

[17] Chau TK, Yu SS, Fernando T, Iu HHC, Small M. (2019) An investigation of the impact of PV penetration and BESS capacity on islanded microgrids–a small-signal based analytical approach. Proceedings of the 2019 IEEE International Conference on Industrial Technology (ICIT), 13-15 February 2019, Melbourne, VIC, Australia.

[18] Anttila S, Döhler JS, Oliveira JG, Boström CJE. (2022) Grid forming inverters: A review of the state of the art of key elements for microgrid operation. Energies, 15(15):5517.

[19] Lamichhane R, Adhikari S. (2023) Virtual Synchronous Generator for Frequency Stability of Microgrid. Proceedings of the International Conference on Role of Energy for Sustainable Social Development; 14th-15th May 2023, Kathmandu, Nepal.

[20] Elwakil MM, El Zoghaby HM, Sharaf SM, Mosa MA. (2023) Adaptive virtual synchronous generator control using optimized bang-bang for Islanded microgrid stability improvement. Protection and Control of Modern Power Systems, 8(4):1-21.

[21] Chakraborty S, Patel S, Salapaka MV. (2022) μ-Synthesis-Based Generalized Robust Framework for Grid-Following and Grid-Forming Inverters. arXiv preprint arXiv:2207.02818. Available: https://arxiv.org/abs/2207.02818

[22] Lin Y, Zhu J, He F. (2025) An improved model-free predictive voltage control for grid-forming inverter with adaptive ultra-local data-model in renewable energy system. Frontiers in Energy Research, 13:1526992.

[23] Miller NW, Sanchez-Gasca JJ, Price WW, Delmerico RW. (2003) Dynamic modeling of GE 1.5 and 3.6 MW wind turbine-generators for stability simulations. Proceedings of the 2003 IEEE Power Engineering Society General Meeting (IEEE Cat No 03CH37491), 13-17 July 2003, Toronto, ON, Canada.

[24] Ngamroo I. (2017) Review of DFIG wind turbine impact on power system dynamic performances. IEEJ Transactions on Electrical Electronic Engineering, 12(3):301-11.

[25] Bossoufi B, Aroussi HA, Ziani E, Lagrioui A, Derouich A. (2014) Low-speed sensorless control of DFIG generators drive for wind turbines system. WSEAS Transactions on Systems and Control, 100:4.

[26] Hieu NH, Luu CT, Tuan TQ. (2015) Effects of DFIG wind power generation on Vietnam power system operation. Proceedings of the 2015 IEEE Eindhoven PowerTech, 29 June 2015 - 02 July 2015, Eindhoven, Netherlands.

[27] Khoa NM, Tung DD, Dai LV. (2022) Experimental study on low voltage ride-through of DFIG-based wind turbine. International Journal of Electrical Electronic Engineering Telecommunications, 11(1):1-11.

[28] Islam H, Mekhilef S, Shah NBM, Soon TK, Seyedmahmousian M, Horan B, Stojcevski A. (2018) Performance evaluation of maximum power point tracking approaches and photovoltaic systems. Energies, 11(2):365.

[29] Gao JT, Shih CH, Lee CW, Lo KY. (2022) An active and reactive power controller for battery energy storage system in microgrids. IEEE Access, 10:10490-9.

https://doi.org/10.31436/iiumej.v27i2.3951

Comparative Performance Analysis of Forward Error Correction Schemes in AX.25-Based APRS Communication Systems with AFSK Modulation using GNU Radio Simulation

RIFKI ARDINAL[1,2], AGFIANTO EKO PUTRA[1*], WAHYUDI HASBI[2], ERIKO NASEMUDIN NASSER[2], WIDYA ROZA[2]

[1] *Department of Computer Science and Electronics, Universitas Gadjah Mada, Yogyakarta, Indonesia.*
[2] *Research Center for Satellite Technology, National Research and Innovation Agency, Bogor, Indonesia*

**Corresponding author: agfi@ugm.ac.id*

(Received: 19 September 2024; Accepted: 10 February 2026; Published online: 10 May 2026)

ABSTRACT: This paper presents a comparative analysis of Forward Error Correction (FEC) schemes in AX.25-based Automatic Packet Reporting System (APRS) communications using AFSK modulation through GNU Radio simulations. The study evaluates uncoded transmission, Reed–Solomon (255,223), several convolutional code configurations, and concatenated Reed–Solomon + convolutional coding under both Additive White Gaussian Noise (AWGN) and Rayleigh fading channels. Results show that convolutional coding with rate 1/3 and constraint length K=7 achieves the best trade-off, reaching error-free performance at Eb/No = -2.5 dB under AWGN and maintaining robustness under fading. Reed–Solomon codes proved effective for correcting burst errors, while concatenated coding showed strong potential but required interleaving for stable operation. These findings highlight that convolutional coding offers the most practical balance between performance and complexity for real-world APRS and satellite systems, providing insights for future enhancements in robust low-SNR communication links.

ABSTRAK: Kajian ini membentangkan analisis perbandingan bagi skim Forward Error Correction (FEC) dalam komunikasi Automatic Packet Reporting System (APRS) berasaskan protokol AX.25 dengan modulasi AFSK melalui simulasi GNU Radio. Penyelidikan ini menilai penghantaran tanpa FEC, kod Reed–Solomon (255,223), beberapa konfigurasi kod konvolusional, serta kod gabungan Reed–Solomon + konvolusional pada saluran Additive White Gaussian Noise (AWGN) dan Rayleigh fading. Dapatan menunjukkan bahawa kod konvolusional dengan kadar 1/3 dan panjang kekangan K=7 memberikan prestasi terbaik, mencapai komunikasi bebas ralat pada Eb/No = –2.5 dB dalam saluran AWGN dan kekal teguh di bawah keadaan fading. Kod Reed–Solomon berkesan untuk pembetulan ralat berurutan, manakala kod gabungan menunjukkan potensi yang kukuh namun memerlukan interleaver untuk operasi yang stabil. Hasil ini menegaskan bahawa kod konvolusional menawarkan keseimbangan paling praktikal antara prestasi dan kompleksiti bagi aplikasi APRS dan sistem satelit, di samping memberikan panduan untuk pembangunan komunikasi yang lebih boleh dipercayai dalam keadaan SNR yang rendah.

KEYWORDS: *APRS, AX.25, AFSK, Forward Error Correction, Reed–Solomon, Convolutional Codes*

https://doi.org/10.31436/iiumej.v27i2.3951

1. INTRODUCTION

The Automatic Packet Reporting System (APRS) is a digital communication system commonly used by amateur radio operators to exchange real-time information, such as location, weather, and short messages [1]. In addition, APRS can be used as a satellite payload for radio communication. The LAPAN-A2 satellite uses APRS and a voice repeater for amateur radio communication missions. Establishing connections between the Indonesian Amateur Radio Community (ORARI) uses amateur radio frequencies for position monitoring, communication from distressed areas, and disaster mitigation through APRS [2]. Satellites have wide coverage that crosses geographic boundaries, provide reliable communication access even in remote areas, and support communication independent of terrestrial infrastructure. [3].

The AX.25 protocol is used for APRS communication. AX.25 is a data-layer protocol designed for amateur packet radio networks. The AX.25 protocol has three types of frames: information frames for data transfer, supervisory frames for transfer acknowledgment, and unnumbered frames for establishing link contact. The use of AX.25 on small satellites can cause significant redundancy and overhead, which is why more efficient protocols with lower redundancy are being developed [4]. As shown in Table 1, an AX.25 frame comprises several main fields, including the flag, address, control, information, and frame check sequence (FCS) fields [5]. AX.25 frames are transmitted using AFSK (Audio Frequency Shift Keying), which is FSK-FM with a 1700 Hz subcarrier and a ±500 Hz deviation on the amateur UHF frequency (437 MHz) [6].

Table 1. Structure of the AX.25 Frame

Flag	Address	Control	Info	FCS	Flag
01111110	112/224 bits	8/16 bits	N*8 bits	16 bits	01111110

The APRS communication system operates with packet data transmitted over amateur VHF/UHF frequencies. Although AX.25, the standard data link layer protocol for amateur packet radio communications, has been shown to produce very low error rates in HF bands [7], when used in the VHF band, communication becomes less reliable and results in a relatively high error rate. This problem directly affects the reliability of data transmission, which is a key performance indicator for APRS utilization. Figure 1 illustrates the overall workflow and infrastructure of APRS. An APRS tracker, or mobile tracker, is a device mounted on a moving object, such as a vehicle or person, that transmits position, message, or status data in an AX.25-based packet format [8].

Forward Error Correction (FEC) techniques are crucial for addressing the problem of noisy channels. FEC is a technique used to detect and correct errors that occur during data transmission by adding redundant data. This is crucial because wireless communication channels are often affected by various noise sources such as white noise, impulse noise, and other interference [9]. With FEC, devices can automatically correct errors, thereby increasing the efficiency and reliability of communication systems without requiring the retransmission of data. Based on the test data, the AX.25 protocol without FEC is susceptible to bit errors and can cause up to 30% packet loss at low signal-to-noise ratio (SNR) conditions (≤10 dB). FEC consists of three types: linear block codes, cyclic codes, and convolutional codes. Linear block codes are a basic form of FEC. This code uses the mathematical structure of vector spaces and subspaces to construct codes that detect and correct errors in transmitted data. Cyclic codes are a subset of linear block codes that have the property that if a codeword is part of the code, then

https://doi.org/10.31436/iiumej.v27i2.3951

every rotation of the codeword is also part of the code. Cyclic codes include BCH and Reed–Solomon codes [10].

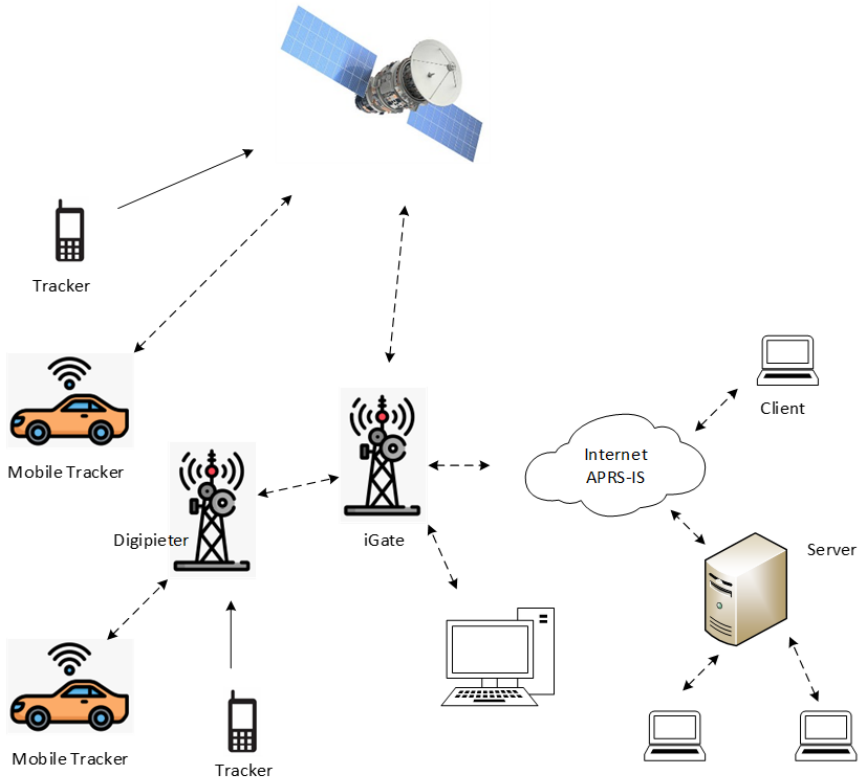


Figure 1. APRS infrastructure

Previous studies generally focused on BCH codes [11] or other more complex FEC codes, such as LDPC and turbo codes [12], but were still limited in their evaluation of Reed–Solomon and convolutional codes in the context of AX.25-based APRS. Furthermore, most performance analyses have been conducted on simple AWGN channels without considering the effects of multipath fading, which is common in satellite communications and dense terrestrial environments. The trade-off among BER performance, complexity, and latency in APRS systems has rarely been discussed. Convolutional codes are widely used in modern communication standards, such as WLAN (802.11ax), LTE, and satellite communications [13]. Meanwhile, Reed-Solomon codes are highly effective block codes for correcting burst errors that often occur in wireless environments [14]. A Reed-Solomon code improvement scheme for distributed storage systems aims to reduce bandwidth requirements. The results show that mathematical optimization can reduce the overhead without compromising the system's error-correcting capabilities of the system [15].

This study aims to evaluate and compare the performance of various Forward Error Correction (FEC) schemes in APRS communication systems based on the AX.25 protocol with Audio Frequency Shift Keying (AFSK) modulation through GNU Radio simulations. The analysis focuses on the Bit Error Rate (BER) performance of several configurations: systems without FEC, systems with Reed-Solomon FEC, systems with convolutional codes in various configurations, and systems with combined coding schemes. The main contribution is the identification of the most practical coding strategy for APRS, demonstrating that a convolutional code with a rate of 1/3 and a constraint length of K=7 offers the best balance of reliability, complexity, and robustness for satellite-based APRS links.

Although several studies have investigated the application of forward error correction in digital communication systems, systematic evaluations of Reed–Solomon and convolutional codes, specifically for AX.25-based APRS links, remain limited, particularly under realistic fading conditions. Moreover, most existing studies focus either on complex modulation schemes or on idealized AWGN channels, leaving a gap in understanding the trade-off between performance, complexity, and robustness when simple, widely adopted modulations, such as AFSK, are employed. This study addresses this gap by providing a comprehensive comparative analysis of practical FEC schemes for APRS in AWGN and Rayleigh-fading environments, drawing on the literature.

2. COMMUNICATION SYSTEM SIMULATION METHODOLOGY

The simulation chain was developed in GNU Radio with custom Python blocks for AX.25 encoding/decoding, AFSK modulation/demodulation, and FEC coding. Figure 2 shows the system-level architecture: AX.25 payloads are encoded, optionally protected with FEC, modulated using AFSK, transmitted over AWGN or Rayleigh-fading channels with calibrated noise, and finally demodulated and decoded before the BER evaluation.

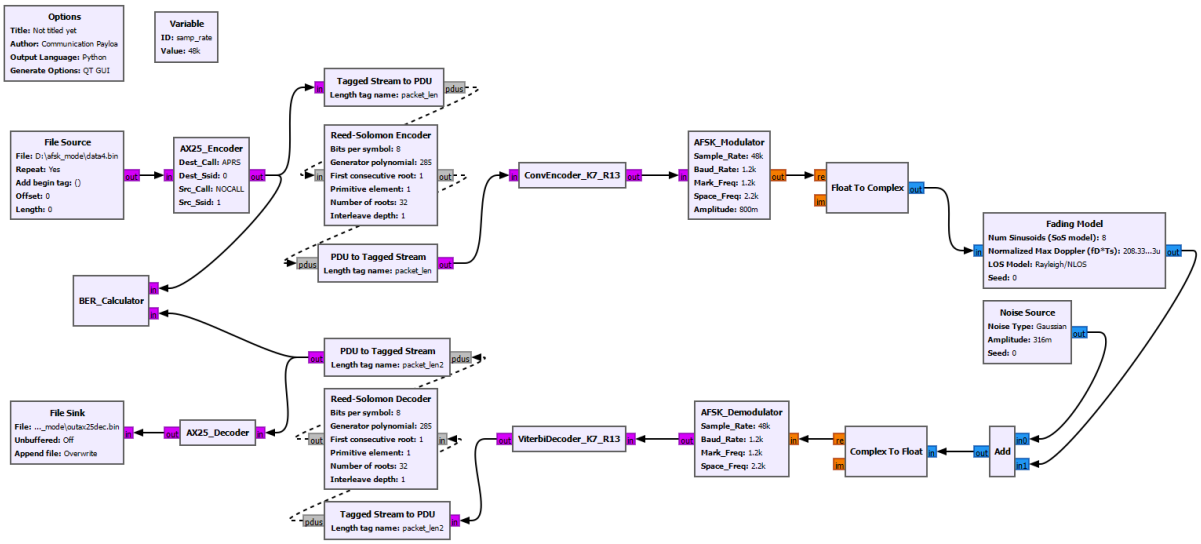


Figure 2. System-Level Simulation Chain with Noise Calibration (σ from Eb/No)

The internal implementation of the AFSK modulator and demodulator follows the standard AX.25 signal processing chain. At the transmitter, NRZI encoding is applied at 1200 bps, followed by tone generation using 1200 Hz (mark) and 2200 Hz (space) sinusoidal oscillators at a sampling rate of 48 kHz. At the receiver, band-pass filtering, tone energy detection, symbol timing recovery, and NRZI decoding are employed to recover the bitstream before AX.25 frame decoding and FEC processing.

This section emphasizes reproducibility and the rationale for parameter selection rather than procedural details. A 16-byte payload was used to represent the typical APRS beacon. AFSK tones at 1200 Hz (mark) and 2200 Hz (space), with a 48 kHz sample rate, were selected in accordance with the AX.25 TNC specification to ensure compatibility with hardware implementations. For block codes, Reed–Solomon (255,223) with a 16-byte correction capability was chosen, reflecting its common use in satellite communications. For convolutional coding, several configurations were tested:

https://doi.org/10.31436/iiumej.v27i2.3951

- Rate 1/2, K = 3, generators [7, 5] (octal)
- Rate 1/2, K = 7, generators [133, 171] (octal)
- Rate 1/3, K = 4, generators [17, 15, 13] (octal)
- Rate 1/3, K = 7, generators [133, 171, 165] (octal)

By default, the configuration R=1/2, K=7, (171,133) corresponds to the CCSDS profile [16]. All the convolutional codes were non-systematic with zero-tail termination. Decoding employed a hard-decision Viterbi algorithm with a traceback depth of ≈ 5×(K–1), in accordance with communication standards. Noise injection was applied to the channel block. The noise standard deviation σ was derived from the target Eb/No using the measured signal power Ps at the modulator's output. This ensured consistent Eb/No calibration across all simulations. Further details of the custom Python blocks, including the implementation of the AFSK modulator/demodulator and AX.25 encoder/decoder, are provided in the Supplementary Material.

Table 2. Key Simulation Parameters

Parameter	Value/Range	Rationale
Payload length	16 bytes	Typical APRS beacon
Modulation	AFSK (mark 1200 Hz, space 2200 Hz)	AX.25/TNC standard
Sample rate	48 kHz	Ensures audio/TNC compatibility
Reed–Solomon code	(255,223), t = 16 bytes	Burst error correction
Convolutional codes	R=1/2, K=3 ([7,5]); R=1/2, K=7 ([133,171]); R=1/3, K=4 ([17,15,13]); R=1/3, K=7 ([133,171,165])	Trade-off analysis
Viterbi traceback depth	≈5×(K–1) (e.g., 30–35 for K=7)	Rule of thumb
Channel models	AWGN; Rayleigh fading (fD/Fs = 10^{-5} – 10^{-3})	Ideal and realistic channels
BER targets	Up to 10^{-3} (reported), extended to 10^{-5}	Satellite benchmark

AWGN was injected at the complex baseband after the AFSK modulator and before the demodulator/decoder, via a linear adder. The average signal power *Ps* was measured at the channel input with noise muted, defined as:

$$Ps = \mathbb{E}\{|s[n]|^2\} \tag{1}$$

For a system with a sampling rate *Fs* and bit rate *Rb*, the standard deviation of the Gaussian noise σ corresponding to a target energy-per-bit-to-noise ratio Eb/No is given by

$$\sigma = \sqrt{\frac{P_s F_s}{2R_b 10^{Eb/No/10}}} \tag{2}$$

This expression follows from Eb = *Ps*/*Rb* and $No = 2\sigma^2$ for complex baseband noise. The computed σ is applied as the amplitude of the Gaussian Noise Source block. For cross-validation, the uncoded AFSK chain was simulated, and the measured BER was compared against the theoretical noncoherent FSK bound.

$$P_b \approx \frac{1}{2} exp\left(-\frac{E_b}{2\,N_o}\right) \tag{3}$$

which provides a practical sanity check to verify the absolute E_b/N_o scaling.

It should be noted that the concatenated Reed–Solomon and convolutional coding scheme implemented in this study does not employ interleaving. Consequently, the observed performance degradation at low Eb/No should not be interpreted as an inherent limitation of concatenated coding itself, but rather as a consequence of the absence of interleaving, which is essential to mitigate burst-error propagation between the inner and outer codes. With appropriate interleaving, more stable performance is expected, particularly in fading channels.

3. PERFORMANCE ANALYSIS AND SIMULATION RESULTS

3.1. BER Measurement Methodology

Bit Error Rate (BER) was used as the primary metric. Noise was injected at the complex baseband after AFSK modulation, with an amplitude σ calibrated to the target Eb/No. BER calculators were placed before and after the FEC encoder and decoder to capture both the raw and corrected error rates. Both AWGN and Rayleigh-fading channels were simulated to represent ideal multipath conditions for satellite/APRS. In this study, BER was measured at two points: pre-FEC (before the encoder/decoder chain) and post-FEC (after decoding). The reported curves focus on the post-FEC BER, which quantifies the residual error rate after error correction, whereas the pre-FEC BER serves as a baseline to characterize channel impairment.

3.2. Comparative Performance on the AWGN Channel

The representative BER values are summarized in Table 3 for Eb/No points ranging from –4 to +4 dB. The convolutional code with a rate of 1/3 and a constraint length of K=7 shows the best performance, achieving an error-free condition at Eb/No -2.5 dB. Meanwhile, Reed-Solomon is proven effective at correcting burst errors but less efficient in fading conditions. The Reed–Solomon (RS) code is widely used in wireless, satellite, digital video broadcasting (DVB), digital audio broadcasting (DAB), digital subscriber line (DSL), and Ethernet communications because it can correct burst and random errors with high coding gain [17].

Table 3. Summary of BER Performance Across Coding Schemes (AWGN Channel)

Eb/No (dB)	Amp	Without FEC	RS (255, 223)	Convolutional Code				RS + Convo
				R=1/3		R=1/2		
				K=7	K=4	K=3	K=7	
-4,0	1,121	1,17E-02	1,17E-02	5,36E-03	5,76E-03	9,25E-03	1,66E-02	5,99E-02
-3,5	1,058	8,89E-03	9,01E-03	2,15E-03	3,14E-03	5,66E-03	9,28E-03	0
-3,0	0,999	6,81E-03	6,64E-03	6,37E-04	2,02E-03	4,07E-03	3,41E-03	0
-2,5	0,943	4,80E-03	4,77E-03	0	1,15E-03	2,25E-03	2,86E-03	0
-2,0	0,890	3,28E-03	3,28E-03	0	5,92E-04	7,90E-04	9,31E-04	0
-1,5	0,840	2,27E-03	1,99E-03	0	1,91E-04	1,17E-03	4,52E-04	0
-1,0	0,793	1,39E-03	1,30E-03	0	6,43E-05	3,43E-04	2,04E-04	0
-0,5	0,749	8,90E-04	6,99E-04	0	5,02E-05	0	1,20E-04	0
0,0	0,707	4,08E-04	3,82E-04	0	2,32E-05	0	0	0
0,5	0,668	1,99E-04	1,58E-04	0	0	0	0	0
1,0	0,630	1,31E-04	7,54E-05	0	0	0	0	0
1,5	0,595	6,08E-05	1,84E-05	0	0	0	0	0
2,0	0,562	2,66E-05	6,81E-06	0	0	0	0	0
2,5	0,530	8,22E-06	5,57E-06	0	0	0	0	0
3,0	0,501	4,31E-06	0	0	0	0	0	0
3,5	0,473	0	0	0	0	0	0	0
4,0	0,446	0	0	0	0	0	0	0

The combined scheme (RS + Convolutional) performs well on AWGN channels at moderate Eb/No. However, its instability at low Eb/No is mainly attributed to the absence of

https://doi.org/10.31436/iiumej.v27i2.3951

an interleaver, which allows burst errors at the output of the inner convolutional decoder to propagate directly into the outer Reed–Solomon (RS) decoder. This behavior should not be regarded as a fundamental weakness of concatenated coding but rather as a limitation of the current implementation. With proper interleaving, improved performance is expected, particularly in fading channels. An interleaver is crucial for improving the iterative decoding performance and reducing the probability of error bursts [18]. The combination of Reed-Solomon and convolutional codes significantly improves BER performance compared with using either code separately [19]. The concatenated code scheme offers an optimal trade-off between complexity, delay, and reliability, making it suitable for communication applications that require high reliability [20].

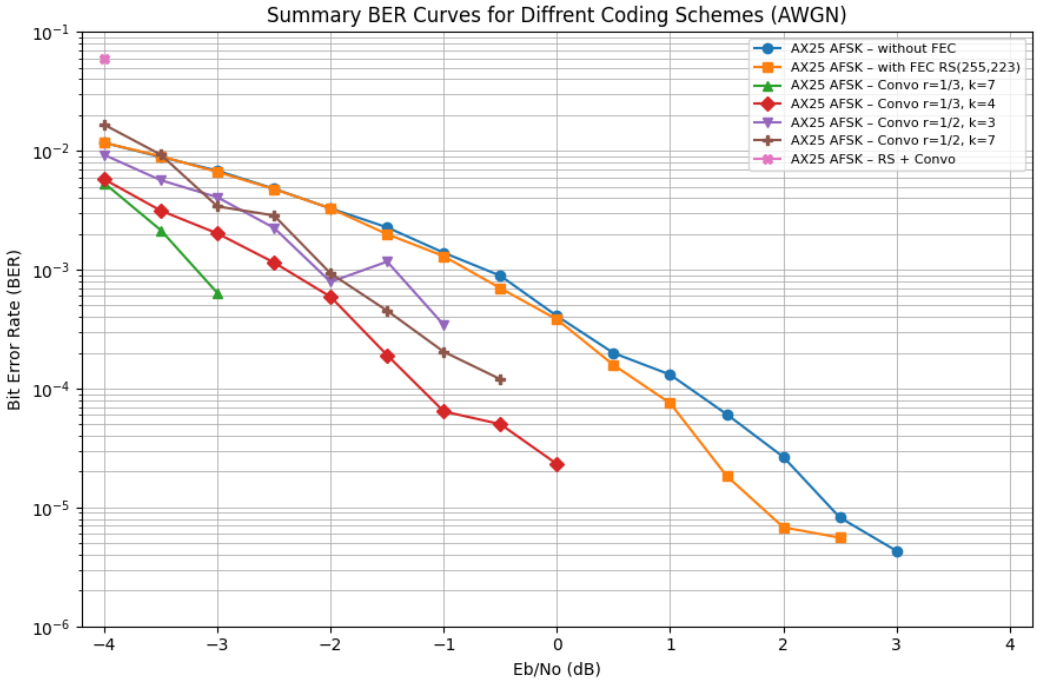


Figure 3. BER vs Eb/No AWGN Channel curve

Figure 3 shows BER curves for different coding schemes under an AWGN channel. The convolutional code R=1/3, K=7 provides the most significant performance improvement, whereas the concatenated (RS+Conv) shows instability at low Eb/No due to the absence of interleaving. It should be noted that the BER performance achieved with AFSK modulation and FEC in this study surpasses the results reported in previous works employing more complex modulation formats [21-22]. In contrast, our results show that convolutional coding (R=1/3, K=7) with AFSK achieves error-free communication at Eb/No ≈ –2.5 dB in the AWGN channel, underscoring the effectiveness of FEC integration even with a relatively simple modulation scheme.

3.3. Performance under Rayleigh Fading

This testing focused on the three most relevant FEC configurations: no FEC, Reed-Solomon FEC (255, 223), and the best convolutional code (rate 1/3, K=7). To simulate the motion conditions, the maximum Doppler value was set based on the vehicle scenario. The data analysis in Table 4 and the curves in Figure 4 show that Rayleigh fading significantly affects the system performance. Compared with testing using AWGN noise alone, the overall BER is much higher at all Eb/No levels, and the curves appear more gradual. Rayleigh fading degrades channel quality relative to AWGN, causing the modulation- or coding-switching

https://doi.org/10.31436/iiumej.v27i2.3951

points to shift to higher SNR values [21]. This confirms that the fading effect causes more significant performance degradation, which cannot be overcome by simply increasing the signal power.

The convolutional code with a rate of 1/3 and K=7 shows the best performance with the lowest BER across the entire Eb/No range. These results demonstrate that robust error-correction mechanisms, such as those implemented by the Viterbi decoder, are more effective at addressing errors caused by fading. This configuration is the only one capable of achieving zero BER at Eb/No of 35 dB and above. The Reed-Solomon scheme also shows a clear performance improvement over the system without FEC, particularly at higher Eb/No values. However, in fading scenarios, its performance is still not as strong as that of the convolutional code.

Table 4. BER Performance Data on the Rayleigh Fading Channel

Eb/No (dB)	Amp	Without FEC	RS (255, 223)	CC (R=1/3, K=7)
-3	0,999	2,97E-02	3,00E-02	3,13E-02
-1	0,793	2,48E-02	2,48E-02	2,54E-02
1	0,630	2,05E-02	2,03E-02	1,66E-02
3	0,501	1,68E-02	1,74E-02	1,31E-02
5	0,398	1,34E-02	1,36E-02	9,21E-03
7	0,316	1,16E-02	1,11E-02	6,32E-03
9	0,251	8,90E-03	9,07E-03	5,02E-03
11	0,199	7,32E-03	7,25E-03	2,61E-03
13	0,158	6,24E-03	5,93E-03	2,48E-03
15	0,126	5,06E-03	4,96E-03	1,46E-03
17	0,100	4,31E-03	4,20E-03	1,25E-03
19	0,079	3,66E-03	3,55E-03	6,31E-04
21	0,063	3,22E-03	3,07E-03	3,24E-04
23	0,050	2,90E-03	2,74E-03	7,90E-04
25	0,040	2,39E-03	2,31E-03	3,52E-04
27	0,032	1,98E-03	2,01E-03	3,99E-04
29	0,025	1,70E-03	1,78E-03	3,27E-04
31	0,020	1,64E-03	1,68E-03	2,16E-04
33	0,016	1,52E-03	1,63E-03	6,75E-05
35	0,013	1,47E-03	1,59E-03	0
37	0,010	1,46E-03	1,56E-03	0
39	0,008	1,44E-03	1,45E-03	0

Notably, the BER of the new convolutional code reaches zero at a very high Eb/No (35 dB), whereas the other configurations never reach it. This indicates that the fading channel model poses a much more challenging problem, in which emerging error patterns can exceed the code's correction capacity even when the signal power is sufficiently high. These results underscore the importance of implementing robust FEC schemes in real-world APRS applications, which typically operate in fading and multipath environments.

https://doi.org/10.31436/iiumej.v27i2.3951

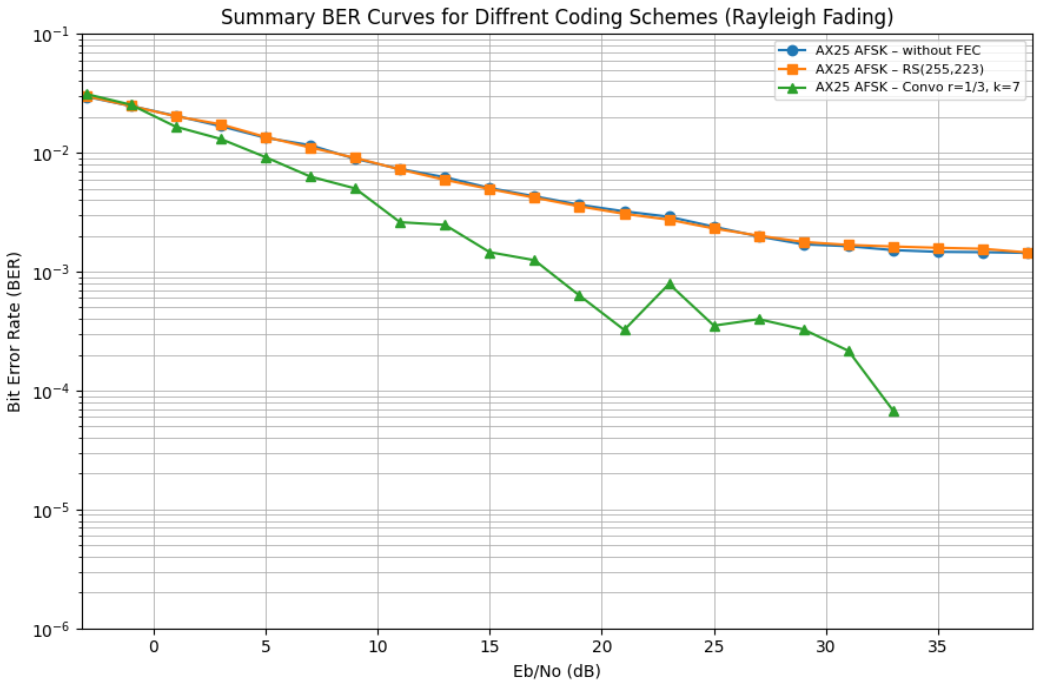


Figure 4. BER vs Eb/No curve - AX.25 AFSK (with Fading Rayleigh)

3.4. BER Floor and Practical Receiver Limitations

The presence of a BER floor in some configurations can be attributed to several factors. First, the use of AFSK modulation with noncoherent detection inherently limits performance at low Eb/No compared to coherent detection schemes. Second, the adoption of hard-decision Viterbi decoding, although computationally efficient, does not fully exploit the soft reliability information of the received symbols, thereby constraining the achievable coding gains. Finally, practical simulation constraints, such as finite traceback depth and synchronization transients, may also contribute to the residual error floor. These observations indicate that further performance improvements can be achieved by incorporating soft-decision decoding and optimized synchronization techniques, particularly for satellite-based and long-range APRS deployments.

4. Implications for System Design

4.1. Performance, Complexity, and Latency Trade-off

Selecting the optimal FEC scheme for a real-world APRS system requires careful consideration of trade-offs among performance, complexity, and latency. The complexity of the Viterbi algorithm increases exponentially with the constraint length of the convolutional code [24]. RS decoding with LCC (Low-Complexity Chase) provides better error correction but is time-consuming; therefore, it is not ideal for high-speed, low-latency applications [25]. Table 5 summarizes the practical trade-offs between the different FEC schemes. A convolutional code with R=1/3 and K=7 offers the best balance for APRS in resource-constrained devices, whereas concatenated codes require significant hardware support.

The choice of the FEC scheme for APRS must balance performance gains against implementation constraints. Convolutional R=1/3, K=7 emerges as the most robust, achieving error-free operation at lower Eb/No while keeping computational complexity within reach of low-power APRS trackers. Reed–Solomon (255,223) is effective in burst-error channels;

however, its block-level decoding introduces latency and memory demands that may be unsuitable for real-time emergency applications. The concatenated (RS+Conv) scheme theoretically offers the best of both worlds, but without interleaving, it becomes unstable at low Eb/No, and its implementation requires FPGA-level resources that many CubeSat payloads lack.

Table 5. Trade-off Analysis of FEC Schemes for APRS System Design

FEC Scheme	Performance (BER)	Complexity	Latency	Resource Requirement
No FEC	Poor (high BER at low Eb/No)	Very Low	None	Minimal (MCU/SDR only)
Reed–Solomon (255,223)	Strong vs burst errors, limited in fading	Moderate (block algebra)	High (block-level decoding)	Requires a memory buffer
Conv. R=1/3, K=7	Best overall, robust in fading	High (Viterbi decoding)	Low (bit-by-bit)	Feasible on DSP/MCU
Concatenated (RS+Conv)	Excellent in AWGN, unstable at low Eb/No without an interleaver	Very High (RS + Viterbi)	High (due to block + interleaver)	FPGA-level resources needed

4.2. Relevance to Real APRS Environment

An analysis of a simple AWGN channel provides a strong initial guide. However, additional testing with a Rayleigh fading channel model provides a more comprehensive picture of the system's robustness in real-world scenarios. This additional analysis with a Rayleigh fading channel is more representative of the APRS environment, which operates at VHF/UHF frequencies and exhibits significant multipath propagation [1-2,21]. The actual APRS environment is much more complex, characterized by fading, non-Gaussian noise, and various types of interference. As shown in the literature, Rayleigh fading is commonly used to model non-line-of-sight channels in satellite and urban communications, resulting in poorer BER performance than AWGN [21]. Tests on a fading channel show that convolutional codes, while very effective at addressing random errors (such as those caused by AWGN noise), also perform reasonably well against fading errors, although their performance is not as good as in AWGN channels. Thus, convolutional codes remain a solid choice for addressing challenging real-world APRS environments.

5. CONCLUSION

This study presented a comparative performance evaluation of several FEC schemes for AX.25-based APRS communication using AFSK modulation in the GNU Radio simulations. The main findings are summarized as follows: First, convolutional coding with rate R = 1/3 and constraint length K = 7 yields the best overall BER performance across both AWGN and Rayleigh fading channels, striking a favorable balance between reliability and implementation complexity. Second, Reed–Solomon coding is effective at correcting burst errors but is less robust in fading environments when used independently. Third, concatenated Reed–Solomon and convolutional coding demonstrates strong potential; however, its instability at low Eb/No in this study is primarily due to the absence of interleaving, rather than an intrinsic weakness of the coding scheme.

These results confirm that, for practical APRS and small-satellite applications, convolutional coding with R=1/3 and K=7 is the most viable solution in terms of performance, latency, and resource constraints. Future work will focus on three main directions: incorporating interleaving into concatenated coding schemes to enhance stability in fading

channels, investigating soft-decision decoding to reduce the observed BER floor, and extending the study to hardware-in-the-loop experiments to validate the computational complexity, latency, and power consumption in real SDR-based APRS implementations.

ACKNOWLEDGEMENT

This work was partially supported by the Department of Computer Science and Electronics, Universitas Gadjah Mada, under the Publication Funding Year 2025. The author would also like to thank the Research Center for Satellite Technology (BRIN) for providing the facilities used in this research.

REFERENCES

[1] APRS Working Group. (2000). Automatic Position Reporting System APRS Protocol Reference Protocol Version 1.0. [Online]. Available: http://www.tapr.org. [accessed: Sept. 2025].

[2] Prahyang SY, et al. (2018). Development of nanosatellite technology with APRS module for disaster mitigation. IOP Conference Series: Earth and Environmental Science, 149(1):012072. https://doi.org/10.1088/1755-1315/149/1/012072.

[3] Elbert BR. (2008). Introduction to Satellite Communication, 3rd ed. Norwood, MA: Artech House.

[4] Akhila G, Sree Vaishnavi R, Poddar PG. (2022). Framing and synchronization of satellite TTC data. In Proc. 3rd Int. Conf. for Emerging Technology (INCET 2022). IEEE. https://doi.org/10.1109/INCET54531.2022.9824011.

[5] Le Roux JH, Barnard MA, Wolhuter R. (2014). Development of a satellite network simulator tool and simulation of AX.25, FX.25 and a hybrid protocol for nano-satellite communications. Stellenbosch University. Available: http://scholar.sun.ac.za.

[6] Busso A, Sousa M, Mascarello M, Stesina F. (2019). A procedure to recover data of a CubeSat mission at very low S/N ratio. In Proc. TTC 2019 – 8th ESA Int. Workshop on Tracking, Telemetry and Command Systems for Space Applications. IEEE. https://doi.org/10.1109/TTC.2019.8895451.

[7] Wicaksono A, Mauludiyanto A, Hendrantoro G. (2020). An HF digital communication system based on software-defined radio. In Proc. ICoSTA 2020. IEEE. https://doi.org/10.1109/ICoSTA48221.2020.1570610561.

[8] Harsono, S. D., Rumadi, & Ardinal, R. (2019). Design and implementation of SatGate/iGate YF1ZQA for APRS on the LAPAN-A2 satellite. IEEE.

[9] Sandoval F, Poitau G, Gagnon F. (2019). Optimizing forward error correction codes for COFDM with reduced PAPR. IEEE Transactions on Communications, 67(7):4605–4619. https://doi.org/10.1109/TCOMM.2019.2910811.

[10] Gazi O. (2019). Forward Error Correction via Channel Coding. Cham: Springer. https://doi.org/10.1007/978-3-030-33380-5.

[11] Wijanto, H., Medina, F. R., Rahmadani, C. P., & Pangestu, B. H. B. (2019). Prototype of telemetry, tracking, and command module for Tel-USat with BCH code

[12] Challa P, Mosa SA. (2018). Performance evaluation and implementation of convolution coded OFDM modem in wireless underwater acoustic communication. International Journal of Communication Systems, 31(14). https://doi.org/10.1002/dac.3737.

[13] Liu, Z., Liu, C., Zhang, H., & Zhao, L. (2025). 76.5-Gb/s Viterbi decoder for convolutional codes on GPU. IEEE Embedded Systems Letters, 17(1), 22–25. https://doi.org/10.1109/LES.2024.3416401.

[14] Yadav, A., Jindal, P., & Basappa, D. (2021). Design and implementation of RS(450,406) decoder: Forward error correction by Reed–Solomon decoding. International Journal of

Embedded and Real-Time Communication Systems, 12(1), 19–43. https://doi.org/10.4018/IJERTCS.20210101.oa2.

[15] Liu, S., Wan, Y., & Xing, C. (2024). Repairing Reed–Solomon codes with less bandwidth. In IEEE International Symposium on Information Theory (ISIT 2024) Proceedings (pp. 494–498). IEEE. https://doi.org/10.1109/ISIT57864.2024.10619655.

[16] CCSDS. (2011). CCSDS 131.0-B-2: Recommendation for Space Data System Standards – TM Synchronization and Channel Coding. CCSDS Secretariat.

[17] Mageswari, N., Mahadevan, K., & Kumar, R. M. (2019). An α-factor architecture for RS decoder implemented on 90 nm CMOS technology for computer computing applications devices. Microprocessors and Microsystems. https://doi.org/10.1016/j.micpro.2019.102855.

[18] Vafi, S. (2021). Parallel concatenated block codes constructed by convolutional interleavers. IEEE Access, 9, 41218–41226. https://doi.org/10.1109/ACCESS.2021.3065236.

[19] Khalil, A., et al. (2019). Combined Reed–Solomon and convolutional codes for IWSN based on IDWPT/DWPT architecture. Procedia Computer Science, 160, 666–671. https://doi.org/10.1016/j.procs.2019.08.095.

[20] Chen J, Chen H, Li Z. (2022). A double serial concatenated code using CRC-aided error correction for highly reliable communication. Computer Networks, 216:109260. https://doi.org/10.1016/j.comnet.2022.109260.

[21] Sarnin, S. S., Kadri, N., Mozi, A. M., Wahab, N. A., & Naim, N. F. (2010). Performance analysis of BPSK and QPSK using error correcting code through AWGN. In Proceedings of the International Conference on Networking and Information Technology. IEEE.

[22] Kaur, S., Singh, N., Kur, G., & Singh, J. (2018). Performance comparison of BPSK, QPSK and 16-QAM modulation schemes in OFDM system using Reed–Solomon codes. In Proceedings of the International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE). IEEE.

[23] Bandiri, S. Y. M., Diakite, S., & Pimenta, T. C. (2023). Analysis of optimum switching points for hybrid adaptive technique under Rayleigh fading channels. In Mixed Design of Integrated Circuits and Systems (MIXDES 2023) (pp. 221–225). https://doi.org/10.23919/MIXDES58562.2023.10203252

[24] Fu, Z., & Liu, H. (2022). Low-complexity hybrid algorithm for decoding convolutional codes. In Proceedings of the 2022 IEEE 8th International Conference on Computer and Communications (ICCC 2022) (pp. 2477–2482). IEEE. https://doi.org/10.1109/ICCC56324.2022.10065728.

[25] Jeong, J., Shin, D., Shin, W., & Park, J. (2021). An even/odd error detection based low-complexity Chase decoding for low-latency RS decoder design. IEEE Communications Letters, 25(5), 1505–1509. https://doi.org/10.1109/LCOMM.2021.3054753.

https://doi.org/10.31436/iiumej.v27i2.3969

Quantitative Assessment of Coffee Quality Using Microwave Non-Destructive Testing

MOHD KHAIRIL ADZHAR MAHMOOD[1,2], ZUHANI ISMAIL KHAN[1*], MEGAT SYAHIRUL AMIN MEGAT ALI[1], AHMAD IHSAN MOHD YASSIN[2], NORZIATON ISMAIL KHAN[3], TAN KWAN YEOW[4]

[1]*Microwave Research Institute, Universiti Teknologi MARA, Shah Alam, Malaysia*
[2]*Faculty of Electrical Engineering, Universiti Teknologi MARA, Shah Alam, Malaysia*
[3]*Faculty of Accountancy, Universiti Teknologi MARA, Shah Alam, Malaysia*
[4]*JWC Roastery Sdn Bhd, Puchong, Malaysia*

**Corresponding author: zuhai629@uitm.edu.my*

(Received: 3 October 2025; Accepted: 22 February 2026; Published online: 10 May 2026)

ABSTRACT: The geographical origin, processing, roast level, and post-roasting resting time affect coffee quality in multiple ways. The traditional quality assessments in the coffee industry (sensory and chemical analyses) are subject to analysts' biases and are destructive, requiring the coffee to be destroyed to run tests. This analysis uses Microwave Non-Destructive Testing (MNDT) to study the dielectric characteristics of specialty coffee (Coffea arabica L.) associated with roasting and resting period. Coffee beans were roasted to produce light, medium, and dark roasts, and the beans were tested for dielectric properties over 30 days in 5-day intervals. The dielectric permittivity was negatively correlated with the level of roasting ($r = -0.590$, $p = 0.005$) and the resting period ($r = -0.483$, $p = 0.027$). The multiple linear regression showed strong predictive ability ($R = 0.762$, Adjusted $R^2 = 0.534$), explaining the effects of roasting and resting period, and their combined effect on the coffee's dielectric properties. This study shows, for the first time, how MNDT can be applied to the coffee industry to evaluate and assess coffee quality.

ABSTRAK: Kualiti kopi dipengaruhi oleh pelbagai faktor seperti asal geografi, teknik pemprosesan, tahap panggang dan tempoh rehat selepas dipanggang. Kaedah tradisional bagi penilaian kualiti, termasuk penilaian deria dan analisis kimia, terdedah kepada subjektiviti dan keperluan ujian yang merosakkan sampel. Kajian ini menggunakan kaedah Microwave Non-Destructive Testing (MNDT) bagi menilai sifat dielektrik kopi istimewa (Coffea arabica L.) yang dipengaruhi oleh tahap pemanggangan dan tempoh rehat selepas dipanggang. Sampel kopi ini dipanggang pada tahap terang, sederhana dan gelap, dengan sifat dielektrik yang diukur selama 30 hari pada selang 5 hari. Keputusan menunjukkan korelasi negatif ketara antara kebolehtelapan dielektrik dan kedua-dua tahap panggang ($r = -0.590$, $p = 0.005$) dan tempoh rehat ($r = -0.483$, $p = 0.027$). Model regresi berganda menunjukkan keupayaan ramalan yang kuat ($R = 0.762$, R^2 Terlaras = 0.534), memberi kesan gabungan pemanggangan dan tempoh rehat pada sifat dielektrik. Penemuan ini menggariskan potensi MNDT sebagai alat berskala, bukan invasif bagi penilaian kualiti kopi, menyumbang kepada amalan pemprosesan lebih baik dan kawalan kualiti dalam industri kopi.

KEYWORDS: *Specialty coffee, Coffee quality, MNDT, Dielectric permittivity, Multiple regression*

https://doi.org/10.31436/iiumej.v27i2.3969

1. INTRODUCTION

Coffee is a widely loved drink worldwide. It has a wide variety of tastes and aromas. The complexity and variety of flavors arise from factors such as where the coffee is grown, how it is processed after harvest, and the specifics of its cultivation and harvest. Coffee is grown all over the world. It is a major world crop. The most widely grown and most traded species is Coffea arabica, or Arabica, which accounts for 57.4% of the world's coffee production. The other 42.6% is produced by Coffea robusta [1]. Coffee that is of "specialty" quality has "uniqueness" as a result of an excellent aroma and flavor. This quality is given to coffee that has received a score of 80 or higher on a 100-point scale from the Specialty Coffee Association (SCA) [2]. Specialty coffee is for consumers with high, specific requirements for aroma, taste, sweetness, acidity, and body [3]. Other factors, like soil, altitude, and processing, are important, but roasting is the most important. Coffee roasting has three levels: light, medium, and dark.

Different roasting times and temperatures produce distinct profiles of aroma compounds and therefore different flavors [4]. Being able to control roasting conditions is crucial for achieving desired flavors. In terms of roasting, the levels of acidity in coffee vary, and more highly roasted beans have lower acidity [5]. Additionally, it was noted that Arabica beans have lower acidity than Robusta beans [6].

Still, examining the coffee after roasting and determining quality is difficult. Evaluating coffee quality has primarily relied on cupping, a subjective sensory analysis, and also includes destructive chemical analyses [7]. Despite widespread consensus that cupping is the best sensory evaluation tool, it is highly subjective and therefore difficult to scale [8]. Chromatography is very thorough but inefficient for large-batch testing and is also considered destructive [9]. Given these issues, it is evident that the specialty coffee market needs better quality control. Dielectric properties without changing the product. An example is evaluating coffee without destroying the sample. To do this, the dielectric properties of the coffee are evaluated, which means the internal properties of the sample. The internal properties are governed by the degree of roasting and the resting period of the coffee beans, which together determine coffee quality. The lack of interest in this area is due to the lack of work on developing the equation that relates the two dielectric characteristics of materials, which is the main hindrance to the many practical uses of MNDT in the coffee industry. Within the coffee industry, among the many dielectric properties, two are of utmost importance, and two are closely related to the key quality parameters of roast coffee and post-roast resting time [10]. MNDT is the first to easily and quickly close the gap between the old analytical methods and the modern approach to the coffee industry.

This study examined the relationships among roast intensity, specialty coffee resting time, and dielectric permittivity over 30 days. The electrical properties of the coffee beans are fundamentally related to the amount of water, their chemical constituents, physical structure, and more, and these characteristics undergo drastic changes when the beans are roasted and then stored. Researchers have examined various aspects of coffee, such as its chemical composition and taste, but little research has examined how these variables affect the electrical properties of coffee using accurate, non-destructive methods.

This study looks at the electrical permittivity of coffee at varying degrees of roasting: light, medium, and dark. It attempts to relate the roasting, aging, and microscopic structural properties of the coffee. Over 30 days, we studied the rapid aging processes of coffee, such as oxidation, loss of carbon dioxide, and moisture movement, which can significantly affect the final quality.

https://doi.org/10.31436/iiumej.v27i2.3969

We believe our study to be the first of its kind in coffee research, as we have, for the first time, combined varying degrees of roasting, different resting time intervals for coffee samples, and measurements of electrical properties using Microwave Non-Destructive Testing (MNDT). Past research on coffee quality has focused on sensory and chemical analyses. There has been little research on the electrical properties of coffee and how MNDT at different degrees of roasting and resting affects them.

This project develops a new technique for determining how roasted coffee beans undergo chemical and physical changes during storage. We demonstrate the first use of microwave-based technologies to monitor these changes in real time. We describe the first use of microwave-based technologies to monitor these changes in real time. We also describe the first use of microwave-based technologies for real-time, non-destructive, and simultaneous evaluation of coffee quality. We also describe the first use of microwave-based technologies to evaluate the quality of coffee and other beverages. Traditional approaches to coffee quality evaluation involve chemical analysis, spectroscopy, and even destructive sampling. The use of MNDT offers the first opportunity to evaluate beans without contact and without destruction during sampling. The multiple regression modeling framework we created strengthens MNDT's predictive analytical power for coffee. The statistical models developed for neural networks identified the critical roasting and aging times that determine the predictive patterns for coffee analysis. The neural networks identified critical patterns in roasting and aging dielectric permittivity and developed a predictive framework for coffee analysis. This methodology aids in implementing quality control systems that provide operational feedback by developing MNCT-based sensors for in-line roasting and real-time coffee monitoring. This research represents the first step in applying MNDT to the coffee industry, with a focus on operational precision. This research provides the industry's first opportunity to use MNDT.

This paper is organized as follows: Section 1 outlines the various methods for evaluating the quality of coffee; Section 2 outlines the proposed methods for research on MNDT; Section 3 discusses the findings of the experiments conducted; and Section 4 discusses the future research and development of the MNDT tools for assessing the quality of coffee.

2. PROPOSED METHOD

The aim of the approach is to employ MNDT as a new tool for assessing coffee quality, specifically during the critical period of resting days after roasting. The present work provides a scientific study of MNDT and demonstrates its novelty and convenience compared with common quality evaluation approaches. Antenna-based free-space MNDT was chosen for non-contact assessment, in contrast to invasive probes, which require contact between the probe and the materials under test.

MNDT uses electromagnetic waves to investigate the interior characteristics of materials without any physical change or damage to the test object. This technology can be used for quality control of coffee to determine critical parameters such as moisture, dielectric properties, and the structure of the bean during resting.

Reflection, transmission, dielectric constant, loss factor, and complex permeability, which depend on temperature and frequency, are assessed by the MNDT method at the microwave spectrum. The principal equations to determine these parameters are as follows [11]:

$$X = \frac{{S_{11}}^2 - {S_{21}}^2 + 1}{2S_{11}} \tag{1}$$

$$\Gamma = \chi \pm \sqrt{\chi^2 - 1} \tag{2}$$

https://doi.org/10.31436/iiumej.v27i2.3969

$$T = \frac{S_{11} + S_{21} - \Gamma}{1 - (S_{11} + S_{21})\Gamma} \quad (3)$$

$$\mu_r = \frac{1 + \Gamma_1}{\wedge(1 - \Gamma)\sqrt{\frac{1}{\lambda_0^2} - \frac{1}{\lambda_c^2}}} \quad (4)$$

$$\varepsilon_r = \frac{\lambda_0^2}{\mu_r}\left(\frac{1}{\lambda_c^2} - \left[\frac{1}{2\pi L}\ln\left(\frac{1}{T}\right)\right]^2\right) \quad (5)$$

In this context, Γ is the reflection coefficient, T is the transmission coefficient, μ_r is the permeability, λ_0 is the free space wavelength, λ_c is the cutoff wavelength, and ε_r is the permittivity.

With proper calibration, the measured values can be related to other material properties, such as defects and moisture content. Free-space measurements are the cornerstone of making MNDT easily accessible for early detection, as they are non-invasive, repeatable, and obtainable in harsh environments, including heat, cold, and complex electromagnetic fields [12].

MNDT is considered in this work as its attributes for which the material is to be evaluated, with no damage in the sample, to sense the value of the dielectric material. The analysis is performed using Microwave Non-Destructive Testing (MNDT) to determine the degree of roasting, rest periods, and the dielectric permittivity of coffee samples. They use a range of X-band frequencies from 8 GHz to 12 GHz, with 3000 data points. First, the MNDT is calibrated with the Through-Reflect-Line (TRL) calibration standards. The measurement is performed on 2 ports (Port 1 and Port 2) because the calculation of dielectric permittivity requires S11 and S21 data. After completing the calibration, the samples are inserted into the acrylic holder, followed by the preparation of the samples above, which were centrally positioned on the spot focusing horn lens. Results were then stored for subsequent analysis. s2p file extensions.

An accurate calibration is important to relate the measured values to known material properties of practical interest, such as flaws and moisture. Calibration ensures the validity and accuracy of measurements by minimizing errors and making results repeatable and comparable across varying conditions. Free-space measurements are non-contact and, if properly calibrated, MNDT methods have the potential to deliver accurate and repeatable results, even under extreme environmental conditions such as heat, cold, and complex electromagnetic backgrounds [13].

This technique eliminates secondary contamination and is advantageous for material analysis. Additionally, MNDT is an important part of materials science and includes a sensor, testing methods, and measurement technologies for detecting damaging cracks and water content in materials using microwave techniques [14]. In recent studies on the conversion of sludge from water-treatment plants into pre-carbonized material, MNDT has been used to assess temperature changes and heating duration [15]. The theory has also been investigated in other disciplines, such as civil engineering, where the material studied has included concrete, and it has applications in food technology. MNDT has been used by researchers to measure the state and characteristics of fruits and animal fat [16].

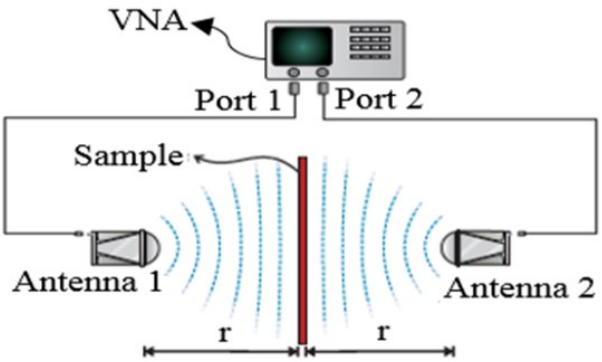
In this work, MNDT is used to assess the object's edibility without directly damaging it, as a convenient method for dielectric characterization of coffee samples. The evaluation is based on the determination of roasting intensity, resting time, and the samples' dielectric permittivity. Measurements are made using X-band (8 - 12 GHz) microwaves with 3000 data points. The X-band (8-12 GHz) offers the best compromise between the ability to penetrate and the ability to resolve in the dielectric measurement of coffee, as the wavelengths within

https://doi.org/10.31436/iiumej.v27i2.3969

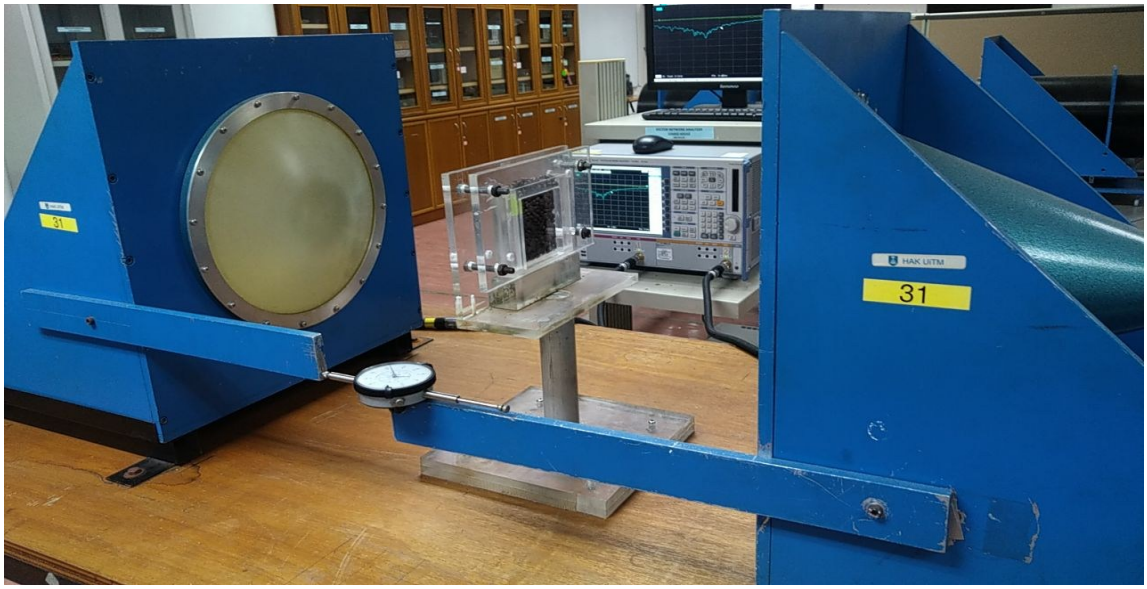
this range are able to cause enough interaction at the bean microstructure level and not too much attenuation, which is a problem in the lower bands, i.e., S-band. S-band is too focused on penetration and bulk material measurement, while X-band is more sensitive to the moisture and structural variations within individual coffee beans [17]. The MNDT system is calibrated using a TRL method prior to measurements, a common experimental approach in microwave measurements to account for systematic errors introduced by the measuring system. The TRL calibration compensates for phase shift, reflection loss, and distortion introduced by the measurement system in the signal path. This process is carried out by the following process:

- *Through Standard*: The system can be calibrated by connecting both ports (Port 1 to Port 2) directly (no sample) to acquire the baseline of the transmission properties.
- *Reflect Standard*: A reflective load (usually a short) is connected to the system to measure its reflection response.
- *Line Standard*: A known-length transmission line is inserted to provide phase information for calibrating the frequency band of interest.

This MNDT setup can accurately determine the scattering parameters (S11 and S21) required for dielectric permittivity calculations once the TRL calibration is complete.



(a)



(b)

Figure 1. (a) MNDT setup (b) Close-up of samples inside the sample holder

https://doi.org/10.31436/iiumej.v27i2.3969

Figure 1 above shows a configuration used in the MNDT. The configuration uses a 2-port Vector Network Analyzer (VNA), with ports 1 and 2 connected to the spot-focused antenna. Spot-focusing horn lens antennas were employed to focus the microwave beam on the sample, mitigating environmental effects in a non-anechoic setup. A sample holder located in the center of the antennas is used to retain the samples to be measured. The calibrated ones are used to cup the coffee samples as per the sample preparation instructions. The holder is placed in the center to take the best signal samples in the interactive spot-focusing horn lens. The S-parameters (S11 for reflection and S21 for transmission) are recorded to assess dielectric properties. The recorded measurements are stored in .s2p format in preparation for analysis.

MNDT is efficient and scalable. It is also the most accurate of the methods discussed. The microwave penetration of the coffee beans' molecular structure enables rapid and accurate scanning of the bean's core. This method is particularly useful during post-roast resting periods, when the coffee undergoes vital chemical and physical changes that enhance flavor. It is only during this period that MNDT can track changes in the bean's core, unlike conventional methods that rely on subjective taste testing. Additionally, precise TRL calibration has been developed to allow for accurate and consistent measurements of MNDT. This allows for reliable assessment of the coffee's roasting level, its electrical permittivity, and the aging period. The ability to measure material properties without damaging the material, coupled with precise calibration, enables MNDT to advance materials science and food technology.

2.1. Research Methodology Overview

Figure 2 describes our methodology used in this research. The flowchart describes the individual components of our computational system. For example, the collection of raw data leads to dielectric property calculations, then to statistical analysis. This provides a solid foundation for our system, allowing us to study how the electrical properties of coffee vary with roasting level and storage duration.

Our data analysis systematizes the following steps: 1. Sample preparation, which includes clarification of raw coffee roast classification, detailed weighing, and MNDT measurement; 2. Import of the recorded data into the MATLAB program; and 3. Execution of the data analysis.

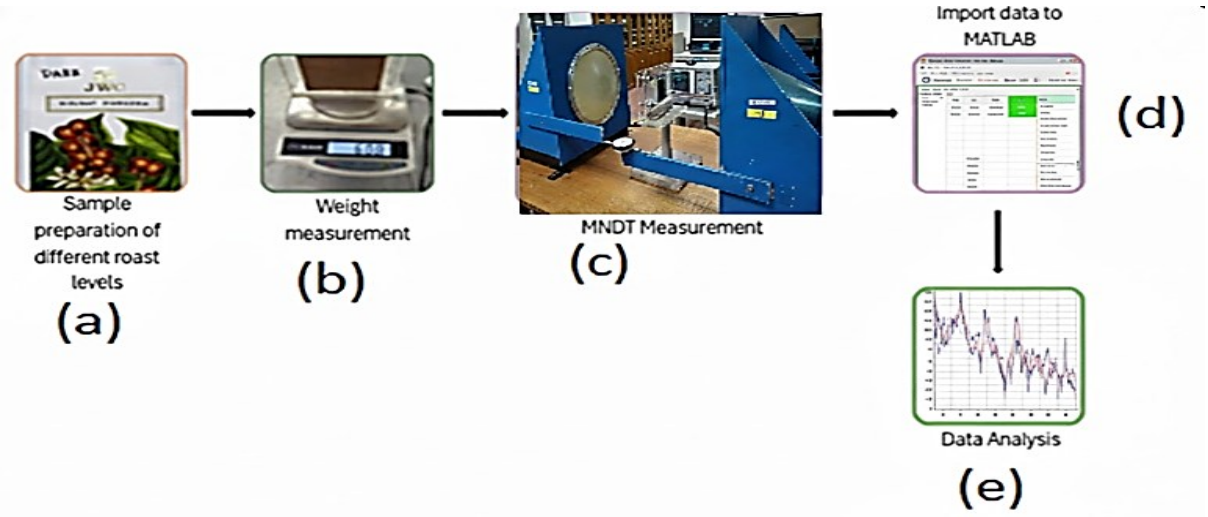


Figure 2. Research methodology overview (a) sample preparation (b) weight measurement (c) MNDT measurement (d) import data to MATLAB (e) data analysis

https://doi.org/10.31436/iiumej.v27i2.3969

2.1.1. Sample preparation

Specialty coffees (Coffea arabica L. Rubiacea) of different roast levels were selected from JWC Roastery (Malaysia) to represent different roasting levels (light, medium, dark) and provide a sufficiently wide base with which to examine the impact of the resting period after roasting.

The coffee samples, kept in their original commercial packaging, were packaged in 200-gram units and were not opened until the day of each measurement. The measurement process was organized over a 30-day period, with data collected every 5 days from day 0 until day 30. This methodology proved to be complete in following up on quality reports and in the dynamic assessment of the coffee.

For each measuring point, a 60 g subsample was taken from the original 200 g packaging and placed in a pre-cleaned acrylic sample holder. A calibrated digital scale was used for weighing. Special attention was paid during preparation to avoid air gaps in the sample holder, which could affect measurement precision. Sample density was controlled by using the same grind size (espresso grind), a fixed-volume acrylic holder, and precise weighing, thereby mitigating density-related skew.

Following these preparations, the coffee sample was measured for MNDT (Figure 1(c)). The experiment was repeated for all 7 samples at the times when the specific roast level was tested. Following the same standardized procedure allows consistent sample prep and handling, and therefore reliable, reproducible measurements comparable across different roast levels and resting periods.

2.1.2. Weight Measurement

To maintain consistency, the sample's accurate weight was ensured using a calibrated digital scale. Samples were weighed under controlled conditions to minimize environmental impact, including temperature and relative humidity. The digital scale was calibrated periodically to ensure accuracy, and all measurements were recorded accurately.

2.1.3. MNDT Measurement

Microwave non-destructive testing (MNDT) was done at X-band frequency with 3000 data points between 8 GHz and 12 GHz. Antennas (spot-focusing horn lenses) were 30 cm apart, with the sample centered. The VNA used is Rohde & Schwarz ZVA40, and the power was set at 0 dBm. Measurements were performed using a calibrated VNA and the two-port free-space measurement technique. Measurement errors were reduced using the Through-Reflect-Line (TRL) calibration to eliminate sources of error that could affect measurements. Each coffee sample was placed in the same acrylic holder to ensure uniform testing conditions. The S-parameters (S11 and S21) obtained were analyzed using the Nicholson-Ross-Weir method to extract the dielectric permittivity. The systematic method used ensures that other researchers can replicate the study.

2.1.4. Data Analysis

The raw S-parameters were imported into MATLAB, where a custom NRW-based script was used to extract dielectric characteristics for each coffee sample. During the preliminary stage of the work, the focus was to extract the critical S11 and S21 values, which are used to determine the reflection and transmission coefficients, respectively. The S12 and S22 parameters were omitted because they were irrelevant to our work on dielectric permittivity.

https://doi.org/10.31436/iiumej.v27i2.3969

Teflon standards operated as our quality control checkpoint. These materials allowed us to validate our NRW calculations for dielectric constants. After we validated our method, we imported the cleaned dataset into SPSS to perform the statistical analysis.

Our analysis was done in two major parts. First, we performed a correlation analysis to assess the relationships among roasting intensity, resting time, and dielectric permittivity. These relationships highlighted the critical parameters of coffee's physical and chemical composition. Second, multiple regression analysis to determine the individual and collective contributions of roast level and coffee age to electrical conductivity. This two-pronged approach demonstrated the complexity of the interplay between these factors in determining coffee's dielectric behavior.

3. RESULTS AND DISCUSSION

Our primary findings focus on the interplay among roasting intensity, aging time, and electrical characteristics. These findings have undergone rigorous statistical analysis and align with the scientific literature.

Our findings provide further insight into the interaction of roasting intensity and post-roast aging on the coffee dielectric values. The greater the roasting intensity, the lower the dielectric permittivity. The decrease in permittivity is attributed to structural changes in the solid matrix and to the amount of moisture present. The dielectric permittivity also decreases, indicating that moisture content and/or chemical changes in the matrix interact with and influence the matrix's electrical composition during the month-long aging process.

3.1. Results

The findings also corroborate earlier studies on the non-destructive assessment of food quality, particularly those using dielectric and spectroscopic techniques. While dielectric techniques are useful for determining some chemical constituents in food (such as acidity and bitterness), their application is usually limited by high cost and the level of expertise required. In such situations, the MNDT technique offers a more viable and effective solution, as it provides a high-quality analysis of the food's bulk properties. Our findings show that MNDT can provide a high-quality analysis of coffee without the need for expensive, highly sophisticated spectroscopic equipment.

Since dielectric permittivity governs energy absorption and transmission, it is an important factor in understanding its role in the interaction between coffee and microwave signals. The coffee samples' dielectric permittivity is shown over a 30-day resting period, illustrated in Figure 2. The results indicate differences in dielectric properties across roast levels (roast level 1: light roast, roast level 2: medium roast, roast level 3: dark roast), suggesting that roast level significantly influences coffee dielectric properties. These properties change over the first 30 days due to moisture redistribution and oxidation, providing a better understanding of the changes coffee undergoes after roasting. The analysis focused on the permittivity (dielectric constant) maximum, which correlated strongly and negatively with roast level and resting period, thereby fulfilling the study's purpose of combining quantitative quality evaluation with MNDT. While the loss tangent was derived from measured S-parameters and is noted as less informative, it is consistent with the literature, in which the dielectric constant is the predominant metric for moisture/structural variation in roasted coffee [18].

In this study, multiple analytical methods are used to ensure a thorough and comprehensive evaluation of the dielectric permittivity measurements. One of these methods, the Nicholson-Ross-Weir (NRW) method, is applied to the raw S-parameter data from the MNDT experiment.

https://doi.org/10.31436/iiumej.v27i2.3969

The NRW method is recognized for its ability to derive values of the complex dielectric constant from microwave measurements involving incident and reflected power. It allows measurement of the dielectric constant with sufficient precision to classify roast levels and track the permittivity of coffee samples over time.

Using data compiled from the first roasting stage (day 0) to 1 month (day 30), the Pearson method was applied to determine the relationship between the dielectric permittivity of coffee and the number of days from roasting. This method analyzes the degree and type of relationship between two continuous variables. The analysis results in a coefficient (r) that ranges from -1 to 1. The closer the number is to 1, the stronger the positive relationship, and the closer it is to -1, the stronger the negative relationship. If r is near zero, that indicates no relationship. The next step was to use multiple regression modeling to evaluate the effect of different independent variables on the dependent variable, as shown in Table 2. This method simultaneously estimates the effects of multiple independent variables on a dependent variable. By doing so, the accuracy of the prediction is improved, making it more precise than that of a one-variable linear predictive model.

3.2. Discussion

Our study examines the effect of different roasting degrees and bean maturation periods on the electrical properties of specialty coffee beans. The application of heat during the roasting process causes a myriad of important chemical reactions and structural changes to the coffee beans, affecting their moisture content, cellular structure, and electromagnetic properties. The subsequent staging process involves several changes, including oxidation, carbon dioxide degassing, and moisture movement, all of which restructure the beans' electrical properties. Such information is important, as it helps coffee professionals and researchers to improve their roasting and bean storage techniques without compromising the quality of the beans.

MNDT is a new tool for coffee evaluation that can be used alongside the established sensory evaluation techniques. The conventional approaches assess the beans through the human senses of smell, taste, and touch, and these assessments vary from individual to individual. Unlike conventional approaches, MNDT provides both a quantitative and an electrical measure, which is guaranteed to yield the same result regardless of who performs the evaluation. This method integrates the art of coffee making with the science of precise measurements and brings to the coffee industry a reliable, scalable quality assurance tool that does not replace the human element but works alongside it.

The speed at which technology is advancing in food science and food material analysis is mirrored by the MNDT system implemented for coffee assessment. The system provides real-time measurement of the dielectric response of aging coffee beans and determines the optimal rest period to achieve the desired taste and product stability. This system can help specialty coffee producers achieve product consistency and reduce the waste from coffee bean storage, thereby increasing the shelf stability of their products. In food science, the use of microwave technology to measure electrical properties and correlate them with chemical changes provides additional tools to answer questions about the chemical changes occurring in food materials during microwave heating.

The commercial prospects are considerable. Automated quality assessment of the industry-level coffee supply chain is possible with MNDT technology. On-farm processing will have more consistent batches and shelf-life forecasting. Bean-to-cup operational streamlining will be achieved, and producers and consumers will have access to the same excellent quality coffee. All coffee supply chain processes will be transformed. The technology can be used to

https://doi.org/10.31436/iiumej.v27i2.3969

quickly and non-invasively measure the degree of roast and the effects of aging on coffee, thereby increasing transparency in specialty coffee. This will create a system of connected knowledge using metrics to inform producers and consumers. We consider the outcomes of this research in the context of industrial innovation and scientific advancement.

3.2.1. Dielectric permittivity

Dielectric analysis has been used in many evaluations of food products. Existing literature has demonstrated that MNDT can be used to measure moisture content and structural changes of different food matrices. This illustrates the technique's reliability and flexibility. Our results also support the reliability of MNDT in assessing coffee quality.

Dielectric analysis has many applications in food systems. This study documents changes in the physical and chemical composition of roasted coffee over time and their impact on MNDT. We document changes in the dielectric properties of roasted coffee samples as a function of roasting level and aging duration. The results show that the dielectric permittivity decreases with higher roasting levels and longer rest times.

Figure 3 illustrates the measurements of dielectric permittivity for coffee samples of varying roast levels (light, medium, dark) and for varying time intervals (Day 0 to 30 days after roasting). All dielectric permittivity values decrease consistently as the roast level changes from light to medium to dark roasts. With light roasts, the dielectric permittivity values are much higher than those recorded for medium plus dark roasts. This phenomenon was consistent across all time intervals from day 0 to day 30, with a negative correlation between coffee roast level and dielectric permittivity. For instance, on day 0 (roasting day), light roast samples recorded a value of 1.14 compared to 1.04 and 1.05 for medium and dark roasts, respectively. This decreasing pattern was consistent across the time intervals. Our results showed consistently decreasing dielectric permittivity values as time advanced across all roast levels. In light roasts, this pattern was notably clear as the dielectric permittivity value plummeted from 1.14 on day 0 to 1.05 by day 25, a decline of almost 0.1 in just 25 days. The findings are consistent with those detailed in [19].

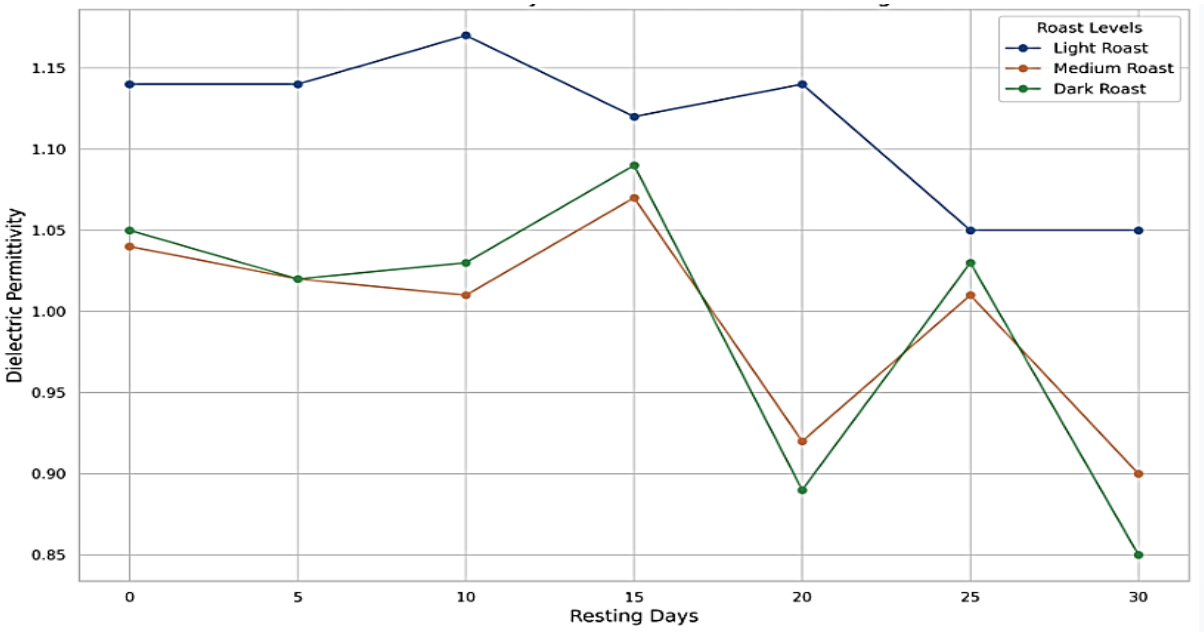


Figure 3. Dielectric permittivity of coffee at different roasting levels

https://doi.org/10.31436/iiumej.v27i2.3969

We observe a decline in dielectric response for both darker roasts and longer resting periods, indicating interactions with roasting intensity and aging. There are two additional factors that contribute to these effects: the degree of roasting and the resting period. This time-related permittivity behavior, consistent across all roast levels, is especially pronounced in medium and dark roast samples, with a steady decrease until day 30.

The decline in permittivity is noticeably more pronounced for dark roasts than for light roasts. Between day 0 and day 25, dark roast samples show a significant reduction (from 1.05 to 0.85), while light roasts exhibit more modest changes (from 1.14 to 1.05). These results show that both the degree of roasting and the time since roasting are important factors in determining the dielectric properties of coffee.

The results show that the combination of dark roasting and longer aging periods for coffee decreases dielectric permittivity values, helping elucidate the physical and chemical changes that occur during roasting and aging. These results corroborate those in [20] regarding the moisture content and its relationship with the roast level: as the roast level increases from light to dark, the moisture content decreases, and the dielectric permittivity and loss tangent values are lower.

Roasting level became the most influential factor impacting dielectric permittivity. Higher roasting temperatures decompose the cellular structure of coffee beans, breaking down their structural matrix and reducing their moisture content. These structural and compositional alterations reduce the coffee beans' dielectric capacity, as indicated by the decreasing permittivity values associated with increased roasting. There was a strong negative correlation between dielectric permittivity and roast level ($r = -0.590$, $p = 0.005$), signifying that roasting reconfigured the physicochemical properties of the coffee beans. The correlation suggests that the alterations from roasting fundamentally influence the dielectric properties of the beans.

In addition to the roast degree, the 30-day resting period was found to be one of the factors causing a decrease in dielectric permittivity. With time, the degassing, moisture loss, and oxidative reactions tend to impair the dielectric properties of coffee beans. The greatest changes in dielectric permittivity were observed for darker roasts, indicating that the degree of roasting increases the effects of aging. The dielectric permittivity and the coffee aging period showed a moderate correlation ($r = -0.483$, $p = 0.027$), indicating that the electrical properties of coffee change as it ages, albeit modestly. This pattern reflects the changes that occur in the beans and the physical and chemical properties of the beans that continue to evolve, and are not limited to roasting.

In summary, the effects of the roasting and aging days on the dielectric properties of the coffee are better understood. The effects illustrate the unique, distinguishing impacts of roast level and aging processes, and how they contribute to the evolution of coffee.

3.2.2. Pearson correlation analysis

This study employs Pearson correlation analysis because it effectively measures the strength and direction of linear relationships between two continuous variables. In this study, the focus was on the relationships among roast levels and resting periods and their effects on dielectric permittivity. Since dielectric permittivity is a quantitative, continuous variable, using Pearson correlation was most effective in measuring the relationship. This method is effective in identifying research trends and guiding future studies.

Table 1 data show a significant inverse relationship between post-roast day storage and coffee dielectric permittivity ($r = -0.483$, $p = 0.027$). This shows that the more post-roast days, the lower the dielectric value, indicating a negative relationship. A correlation coefficient of -

0.483 is interpreted as a moderate inverse linear correlation. This may result from the loss of volatile compounds, fluctuations in moisture, and chemical reactions that occur during storage. The first 10 days post-roasting showed the most significant reduction in permittivity, after which it stabilized. This means that during the initial days of storage, the coffee undergoes the most significant changes, such as degassing and oxidative reactions. After that initial period, the rate of dielectric drop declines, indicating little to no change, which implies the chemical reactions have reached equilibrium.

Table 1. Pearson correlation analysis

Variable 1	Variable 2	Pearson Correlation	Significance (2-tailed)	N
Days	Roast	0	1	21
Days	Dielectric Permittivity	-0.483	0.027	21
Roast	-0.59	-0.59	0.005	21

The correlation between the level of roast and the dielectric permittivity of coffee is statistically significant and negative ($r = -0.590$, $p = 0.005$). Thus, the higher the roast level of coffee, the lower the coffee dielectric permittivity. This can most likely be attributed to the destruction of some cellular structures and the loss of moisture 'trapped' within them during roasting at very high temperatures. Because the moisture content in cellular structures is a significant factor affecting the dielectric properties, loss of moisture in the darker roasts contributes to a decrease in dielectric permittivity. A correlation coefficient of -0.590 represents a strong negative linear relationship between the two variables. The two negative correlations are significant at the 0.05 and 0.01 levels. Hence, there is a definite strong relationship between the roast level and the dielectric permittivity of coffee, and lower roast levels will correspond to higher dielectric permittivity and tangent loss. Roasting coffee leads to changes in its chemical composition, caused by the breakdown of polysaccharides, proteins, and lipids. These breakdown products of organic compounds cause a change in the material to behave in a way that the dielectric permittivity is lower than before. The darker the roast, the more intense the heat, and the greater the loss of the dielectric properties.

These findings provide a solid foundation for understanding coffee's dielectric properties in relation to roasting and aging processes. The inverse relationships/correlations show the declining dielectric permittivity of coffee as the roasting and aging processes continue to alter the beans. This information is a starting point for understanding how roasting processes can be improved, the changes in a coffee's electrical properties during storage, and the development of new methods to assess coffee quality non-destructively using the dielectric properties of a coffee sample.

The numerous relationships presented here do not imply a cause-and-effect pairing, and therefore, exploratory research would be required to identify the cause-and-effect relationships and determine the mechanisms involved. Sample size, a lack of control for confounding variables, uncontrolled variables, measurement precision, and measurement bias may have a significant effect on the correlation and the strength and significance of the relationship

The correlations among aging period, roasting level, and dielectric permittivity provide insight into the relationships involved, and ultimately demonstrate the value of dielectric measurement in understanding, controlling, and improving the processes associated with coffee and its storage.

3.2.3. Multiple regression analysis

The multiple regression analysis in Table 2 builds on previous Pearson correlation findings and aims to develop a more robust model to explain variation in the dielectric permittivity of coffee. These relationships are vital for improving roasting methods, maintaining freshness and quality consistency (enabling non-destructive quality assessment), and accounting for variations in the coffee's dielectric properties. In this analysis, both post-roast aging (days) and roasting degree (intensity) are included as independent variables and are quantified individually and in combination, helping explain variations in the dielectric properties of coffee.

Table 2. Multiple regression analysis

Metric	Value
R	0.762
R-square	0.338
Adjusted R-square	0.314
Std. Error of the estimate	0.0605
F	12.462
Resting days	-0.483
Roast level	-0.59

Table 2 presents the regression analysis of dielectric permittivity as the dependent variable, with post-roast aging (Days) and roasting degree (Roast) as the independent variables. The regression analysis summary indicates a multiple correlation of R = 0.762, indicating a strong linear relationship with the model's independent variables. It is also noted that the adjusted R-squared is 0.534, implying that the combination of the independent variables, post-roast aging and roasting degree, explains 53.4% of the variation in the dielectric permittivity of coffee.

The ANOVA table shows that the model is significant ($F = 12.462$, $p < 0.001$), indicating that at least one predictor is significant in predicting the dielectric permittivity. The regression coefficients are explained more by the predictor variables. The beta value represents the strength of the relationship between the dielectric permittivity of coffee, the number of days since roasting (resting period), and the roast level.

- *Resting period (Days)*: The standardized coefficient of -0.483 has a p-value of 0.005, with a significant level at $p<0.01$, suggesting that an increase in the number of days since roasting (resting period) is associated with a significant decrease in dielectric permittivity, after controlling for the effect of roast level.
- *Roast level (Roast)*: The standardized coefficient of -0.590 has a p-value of 0.001, with a significant level at $p< 0.01$, indicating that an increase in the roast level is associated with a significant decrease in dielectric permittivity, after controlling for the effect of the number of days since roasting (resting period).

Both predictor variables (Number of Days and Roast) significantly contribute to the regression model, with the roast level having a slightly stronger negative association with dielectric permittivity than the number of days since roasting (resting period). Therefore, in general, it can be inferred that the dielectric permittivity of coffee is strongly dependent on resting period and roast level, with the darker roasts and longer resting period having lower dielectric permittivity than the lighter roasts. With these two predictors, the multiple regression model explains a large fraction of the variation in dielectric permittivity, indicating their usefulness for modeling and predicting changes in the dielectric properties of coffee during roasting and aging.

The results of multiple regression analyses indicate that both roasting and aging periods have independent, significant contributions to the variance explained, and that their combined effect explains a significant portion of the variance (Adjusted $R^2 = 0.534$). With the p-value under 0.01, the independent yet inclusive characteristics of the independent variables confirm the predictors' roles in the complexity of coffee quality.

The most significant predictor was roasting. This is attributed to the structural and compositional changes induced by roasting. Multiple studies have shown that coffee roasting induces changes to a coffee bean's moisture content, its cellular makeup, and the carbonation that develops during roasting. These changes also affect the beans' coffee's electromagnetic response and explain why the roast level is one of the most significant predictors of the coffee's dielectric response.

The resting period was the weakest predictor, but it was still significant in the model. This indicates that this period is important and should not be neglected during coffee post-roast processes if good-quality coffee is to be achieved and maintained. The resting period is the time during which volatile compounds equilibrate, and pressure is allowed to equilibrate (internal and external pressures equalize). During the resting period, dielectric permittivity values may change significantly.

Our research emphasizes the significance of tracking resting duration across study or business operational data collection. This tracking establishes consistent benchmarks for the quality of the coffee beans. The relation of roasting and post-roast aging strongly influences coffee's dielectric response and illustrates the integration of roasting and post-roast aging into quality frameworks.

There are strong statistical grounds for the model, validating further work on the many relationships arising from the interplay between processing variables and coffee attributes.

3.2.4. Research contribution

Our findings align with existing literature on the dielectric properties of food systems, particularly the studies on the effects of moisture on the permittivity of foodstuffs. Coffee exhibits hygroscopic behavior, leading to large differences depending on storage temperature and environmental conditions, and the identified trends suggest that dielectric permittivity may serve as a useful indicator of the degree of freshness and quality of coffee. This study builds on existing knowledge of the effects of roasting and aging on coffee quality indices.

From a commercial perspective, the study offers a time frame. In the first two weeks post-roasting, dielectric properties should reflect changes in coffee freshness and quality. In coffee science, this study also enhances the application of MNDT. The findings show that, compared with established protocols, the method can better support sensory evaluations that involve subjectivity and variability. MNDT provides objective and measurable metrics of the desired quality attributes, which may facilitate the first degree of automation in the coffee value chain.

Our research demonstrates that MNDT has significant potential as an innovative technology for understanding and assessing coffee quality. These findings also demonstrate that MNDT is reliable, scalable, and non-destructive for evaluating the quality of post-roasted coffee in commercial roasting facilities. Measuring changes in dielectric properties would help coffee roasters and researchers obtain precise, quantifiable metrics to assess optimal roasting profiles and storage durations. The proposed solution has the potential to be integrated into the coffee roasting and packaging industry through highly automated, high-speed systems that align with and meet industry needs.

Our research also illustrates the importance and necessity of understanding and evaluating the roasting intensity and resting time of coffee beans in order to develop and implement dielectric-based quality assessment methods. MNDT technology can serve as a quality control instrument in the coffee roasting industry to address a wider range of problems, such as optimizing the storage of roasted coffee beans. This will help provide product quality uniformity in different batches.

Rather than relying on subjective sensory analysis, which can vary by evaluator, the MNDT method uses objective measurement, thereby enhancing its objectivity. MNDT provides a comprehensive service, whereas the other methods provide qualitative analysis. MNDT is also the most practical method for operational monitoring of quality control measures, as it significantly reduces the need for destructive testing, lowers manpower requirements, lowers operational costs, and enhances the preservation of coffee samples. The combination of these practical aspects makes MNDT a very innovative system for the coffee quality control industry.

3.2.5. Study Limitations and Future Directions

This study offers the first look into the dielectric properties of roasted coffee. Additional studies will be needed to consider the numerous other factors that affect dielectric permittivity. Future studies will include other coffee varieties (e.g., Coffea robusta) and different roasting and processing profiles to broaden the generalizability and validate the proposed method.

There is also a need for further mechanistic studies to determine the chemical and structural factors that differ within coffee and lead to different dielectric properties. This would assist in understanding the aging of coffee, the roasting process, and the underlying factors driving changes in coffee quality.

The proposed method has been developed without considering real-world industrial applications, and as such, it has been implemented with multiple limitations. There is a need for further studies that demonstrate real-world applications of the proposed method, show real-time tracking and monitoring during roasting and storage, and enable its full potential in the coffee industry. This will allow further refinements to the proposed method and simplify its use by farmers to determine the quality of harvested coffee. These findings continue to demonstrate the potential of dielectric permittivity as an indicator of coffee quality, solidifying the case for the continued development of dielectric permittivity-based quality control methods and for improving the quality of coffee specialty production.

4. CONCLUSION

The research has shown that roast level and time since roast (resting period) affect the dielectric permittivity of the coffee samples. Correlation analysis and the multiple regression model have shown strong negative relationships for the predictor variables against the dielectric permittivity measurements. In regression analysis, it explained around 58.1% of the variation in dielectric permittivity, which may lend some validity to considering it a predictor. The findings seem to affirm that the lower dielectric permittivity values of dark roasts and longer resting periods are due to different chemical/structural aspects of the coffee that change during roasting and resting. Mechanisms underlying the changing relationships and dielectric permittivity values for potential non-destructive quality control in the coffee industry need to be investigated further. In addition, other factors, such as the variety of beans, growing conditions, and bean processing, may need to be studied to assess their dielectric properties. Ultimately, measuring dielectric permittivity and integrating it into existing quality control

systems may simplify and enhance the sustainability and profitability of the coffee supply chain, including processing, storage, and quality assessment.

ACKNOWLEDGEMENT

The author would like to thank JWC Roastery Sdn Bhd and Microwave Research Institute for providing the infrastructure, assistance, and sample provisions.

REFERENCES

[1] E. R. Guimarães, P. H. M. V. Leme, D. C. De Rezende, S. P. Pereira, and A. C. Dos Santos, "The brand new Brazilian specialty coffee market," *Journal of Food Products Marketing*, vol. 25, no. 1, pp. 49–71, Jan. 2019, doi: 10.1080/10454446.2018.1478757.

[2] A System to Assess Coffee Value: Understanding the SCA's Coffee Value Assessment [https://static1.squarespace.com/static/584f6bbef5e23149e5522201/t/667182ffdde8a5081afc2d8c/1718715138872/SCA+-+A+System+to+Asssess+Coffee+Value+-+June+2024+(Secured).pdf]

[3] D. A. Campo-Ceballos and C. A. Gaviria-López, "Optimising consumer acceptability of Cauca specialty coffee through roasting profiles on acidity and body sensory attributes," *Czech Journal of Food Sciences*, vol. 41, no. 1, pp. 64–72, 2023, doi: 10.17221/181/2022-CJFS.

[4] M. Gancarz *et al.*, "Impact of coffee bean roasting on the content of pyridines determined by analysis of volatile organic compounds," *Molecules*, vol. 27, no. 5, Mar. 2022, doi: 10.3390/molecules27051559.

[5] E. B. Tarigan, E. Wardiana, Y. S. Hilmi, and N. A. Komarudin, "The changes in chemical properties of coffee during roasting: A review," in *IOP Conference Series: Earth and Environmental Science*, IOP Publishing Ltd, Feb. 2022, doi: 10.1088/1755-1315/974/1/012115.

[6] S. E. Yeager, M. E. Batali, J. X. Guinard, and W. D. Ristenpart, "Acids in coffee: A review of sensory measurements and meta-analysis of chemical composition," *Crit. Rev. Food Sci. Nutr.*, vol. 63, no. 8, pp. 1010–1036, 2023, doi: 10.1080/10408398.2021.1957767.

[7] S. D. Rodrigues *et al.*, "Sensory Q-grader evaluation of fermented Arabica coffees by yeast (*Saccharomyces cerevisiae*) and lactic bacteria (*Pediococcus acidilactici*) cultures," *Coffee Sci.*, vol. 15, no. 1, pp. 1–9, 2020, doi: 10.25186/.v15i.1690.

[8] L. L. Pereira, W. S. Cardoso, R. C. Guarçoni, A. F. A. da Fonseca, T. R. Moreira, and C. S. ten Caten, "The consistency in the sensory analysis of coffees using Q-graders," *Eur. Food Res. Technol.*, vol. 243, no. 9, pp. 1545–1554, Sep. 2017, doi: 10.1007/s00217-017-2863-9.

[9] S. Eticha and T. Bedassa, "Determination of caffeine in coffee samples by high performance liquid chromatography and ultra violet-visible spectrophotometry methods from Wollega, Ethiopia," *Int. J. Biochem. Biophys. Mol. Biol.*, vol. 5, no. 1, p. 8, 2020, doi: 10.11648/j.ijbbmb.20200501.12.

[10] A. Salim *et al.*, "Classification of coffee roast levels using multi-layer perceptron on MNDT S-parameters," in *Proc. IEEE Symp. Wireless Technol. Appl. (ISWTA)*, 2022, pp. 88–92, doi: 10.1109/ISWTA55313.2022.9942782.

[11] A. M. Nicolson and G. F. Ross, "Measurement of the intrinsic properties of materials by time-domain techniques," *IEEE Trans. Instrum. Meas.*, vol. 19, no. 4, Nov. 1970.

[12] W. Herr, Ed., *CAS–CERN Accelerator School: Advanced Accelerator Physics*. Geneva, Switzerland: CERN, 2014, doi: 10.5170/CERN-2014-009.

[13] M. A. Farag, A. Zayed, I. E. Sallam, A. Abdelwareth, and L. A. Wessjohann, "Metabolomics-based approach for coffee beverage improvement in the context of processing, brewing methods, and quality attributes," *Foods*, vol. 11, no. 6, p. 864, Mar. 2022, doi: 10.3390/foods11060864.

[14] Y. Sun, X. Wang, C. Zhang, and M. Zuo, "Multiple regression: Methodology and applications," in *Highlights in Science, Engineering and Technology*, 2023, pp. 542–548.

[15] M. A. H. Zable *et al.*, "Investigation on dielectric properties of sludge waste from water treatment using microwave non-destructive testing (MNDT)," *Int. J. Integr. Eng.*, vol. 15, no. 3, pp. 163–169, 2023, doi: 10.30880/ijie.2023.15.03.016.

[16] A. L. W. Awang *et al.*, "Intelligent method for classifying the freshness of coffee," in *Proc. IEEE Symp. Wireless Technol. Appl. (ISWTA)*, 2023, pp. 115–122, doi: 10.1109/ISWTA58588.2023.10249522

[17] R. Tao, L. Xu, X. Bai, and J. Chen, "Principles, applications, and future evolution of agricultural nondestructive testing based on microwaves," *Sensors*, vol. 25, p. 4783, 2025, doi: 10.3390/s25154783.

[18] R. Torrealba-Meléndez, M. E. Sosa-Morales, J. L. Olvera-Cervantes, and A. Corona-Chávez, "Dielectric properties of beans at different temperatures and moisture content in the microwave range," *Int. J. Food Prop.*, vol. 19, no. 3, pp. 564–577, 2016, doi: 10.1080/10942912.2015.1038565.

[19] N. Makul, "Dielectric permittivity of various cement-based materials during the first 24 hours hydration," *Open J. Inorg. Non-Met. Mater.*, vol. 03, no. 04, pp. 53–57, 2013, doi: 10.4236/ojinm.2013.34009.

[20] P. D. Chauhan, D. H. Gadani, and V. A. Rana, "Effect of moisture content variation on dielectric properties of various plant leaves at microwave frequencies," *Sci. Rep.*, vol. 14, no. 1, Jun. 2024, doi: 10.1038/s41598-024-64266-3.

https://doi.org/10.31436/iiumej.v27i2.3988

Experimental Validation of a LabVIEW-Based SFCW Software-Defined Radar System

NOR HAZIMA ARDZEMI, FARAH NADIA MOHD ISA, NORUN FARIHAH ABDUL MALEK*, HUDA ADIBAH MOHD RAMLI, AHMAD FADZIL ISMAIL

Department of Electrical & Computer Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia

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

(Received: 15 October 2025; Accepted: 24 April 2026; Published online: 10 May 2026)

ABSTRACT: Software-defined radar (SDR) provides a flexible and cost-effective platform for implementing and testing diverse radar techniques without hardware redesign. Among these techniques, Stepped-Frequency Continuous Wave (SFCW) radar has shown promise in addressing some of the limitations of Frequency Modulated Continuous Wave (FMCW). This paper presents the experimental validation of an SDR system based on the SFCW technique. The radar system was developed using a USRP N210 with a WBX daughterboard, fully controlled through LabVIEW for waveform generation, data acquisition, and signal processing. Experimental validation was conducted in a controlled indoor environment using a 30 × 30 cm metal target placed at distances from 0.7 m to 1.25 m. According to the data, this research has shown accurate detection at longer ranges, with error rates below 0.5%. However, the increasing inaccuracy at lower ranges highlighted issues in near-field measurements. The findings have confirmed the feasibility of SFCW-based SDR for compact, reconfigurable radar platforms and suggest opportunities for enhancing near-field accuracy through advanced signal processing and calibration.

ABSTRAK: Perisian takrifan radar (SDR) menyediakan platform fleksibel dan kos efektif bagi melaksana dan menguji pelbagai teknik radar tanpa reka bentuk semula perkakas. Antara teknik ini, radar gelombang berterusan frekuensi lanjutan (SFCW) telah menunjukkan kebolehan dalam menangani beberapa batasan gelombang berterusan modul berfrekuensi (FMCW). Kajian ini membentangkan pengesahan eksperimen sistem SDR berdasarkan teknik SFCW. Sistem radar dibangunkan menggunakan USRP N210 bersama papan WBX. Ia dikawal sepenuhnya melalui LabVIEW bagi penjanaan bentuk gelombang, perolehan data dan prosesan isyarat. Pengesahan eksperimen dijalankan dalam persekitaran dalaman terkawal menggunakan sasaran logam 30 × 30 cm yang diletakkan pada jarak 0.7 m hingga 1.25 m. Dapatan penyelidikan ini menunjukkan pengesanan tepat pada julat lebih panjang, dengan kadar ralat di bawah 0.5%. Walau bagaimanapun, ketidaktepatan semakin meningkat pada julat lebih rendah, menyerlahkan isu pengukuran medan. Penemuan ini mengesahkan kebolehlaksanaan SDR berasaskan SFCW pada platform radar yang padat dan boleh dikonfigurasikan semula. Cadangan bagi meningkatkan ketepatan medan adalah melalui pemprosesan dan penentu ukuran isyarat lanjutan.

KEYWORDS: *Software-Defined Radar (SDR), Stepped Frequency Continuous Wave (SFCW), USRP, LabVIEW, Range Detection.*

https://doi.org/10.31436/iiumej.v27i2.3988

1. INTRODUCTION

Radar systems are being used widely in sensing applications such as airport surveillance, landslide monitoring, mining, security screening, and health screening. Fig. 1 illustrates how radar operates in sensing applications. Radar operates by transmitting electromagnetic waves. When the waves hit objects, they will be reflected to the radar. The radar receives reflected waves and analyzes them to determine the characteristics of objects, such as size, velocity, and distance. Over time, technological advancements have led to SDR, an evolution of conventional radar systems. While the principles remain the same, SDR differs in how the key functions are implemented. Instead of relying on fixed hardware circuits, SDR shifts key operations, such as waveform generation, signal control, and signal processing, to software. This software-driven approach not only makes SDR highly flexible for testing but also reduces hardware dependency, lowers costs, and allows rapid reconfiguration across different frequency bands and processing techniques [1].

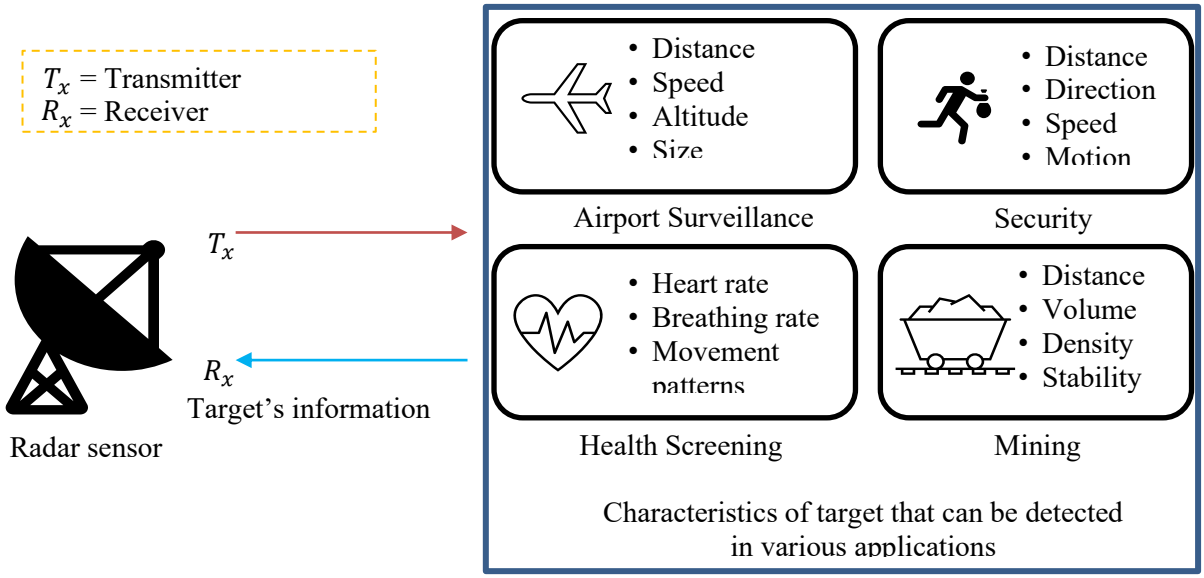


Figure 1. Concept of radars

Among the many radar techniques that can be implemented on SDR platforms, FMCW and SFCW are two of the most widely used. Both approaches can deliver high-resolution, accurate target detection, but they differ in their trade-offs, limitations, and suitability for wideband applications.

FMCW has been widely adopted in SDR research for both near-field and far-field sensing. For example, recent research by [2] has analyzed the performance of an FMCW SDR for maximum-range detection. The research has demonstrated that increasing sweep time extends the maximum detection range in FMCW-based SDR. However, according to [3], for a wide-bandwidth application, increasing the sweep time will result in an inconsistent frequency-sweep rate. On top of that, research in [3] further emphasized that such inconsistencies may degrade detection accuracy. The limitations of FMCW techniques were further identified by [4], who found that power feedthrough was a major drawback, as strong echoes are detected at the receiving antenna due to continuous signal transmission and reception. Beyond these hardware constraints, FMCW techniques often require complex signal processing to mitigate synchronization errors and random delays introduced by SDR communication protocols, as

mentioned in [5]. For instance, a recent work on range estimation in [6] has explored the use of unsynchronized FMCW receivers on LimeSDR platforms. The research approach required a complex setup with a direct signal reference between antennas to compute range and was limited to a 58 MHz bandwidth. The findings indicate an average relative error of 2.15% to 4.28%. Overall, these limitations, inconsistent frequency sweeping rate, power feedthrough, and high signal processing overhead make FMCW less suitable for applications that demand both high resolution and reliable detection accuracy.

In comparison, SFCW offers a solution to the power feedthrough issue while providing a stable frequency reference. According to [7] and [8], SFCW has clear advantages over FMCW, particularly in accuracy and resolution. Its potential has already been validated in diverse fields, for example, in hydrological sensing, where [9] employed an SFCW-based SDR using a USRP X300 over a wide 150 MHz to 6 GHz bandwidth to penetrate complex, multi-layered snowpacks and accurately measure both snow water equivalent (SWE) and snow depth. Another field that employs the SFCW radar system is biomedical research, where [10] has demonstrated that SFCW SDR platforms using the BladeRF achieved sufficient resolution for microwave breast cancer imaging, further confirming the technique's potential for precision detection tasks.

Despite these advances in SFCW radar, a clear research gap remains: existing SFCW studies are predominantly application-specific and often rely on complex, high-cost system configurations. For example, [7] employed the high-cost USRP X310 with a complex multi-tone synchronization architecture, which was validated only through loopback testing, without actual target detection in an open environment. Meanwhile, [8] used a custom modular hardware platform that offers limited software reconfigurability. In addition, the system developed in [9] required an 80-second sweep time per scan due to its wide-frequency-range operation, making it unsuitable for rapid-prototyping scenarios. On the other hand, FMCW-based SDR implementations still face fundamental hardware and signal-processing limitations, as discussed above. Consequently, the development of compact, low-cost, and general-purpose SFCW SDR platforms that integrate hardware and software into a unified framework for rapid prototyping and multirange experimental validation remains limited.

To address the gaps mentioned, this research presents the development of a low-cost SFCW radar system using a USRP N210 equipped with a WBX daughterboard as the hardware platform and fully integrated with LabVIEW. The proposed system leverages discrete frequency stepping to avoid synchronization complexities and hardware overhead while ensuring high-precision detection without requiring complex reference signals or specialized delay-mitigation algorithms. The proposed system also offers flexible waveform generation, real-time hardware configuration, and integrated signal processing within a unified framework, thereby enabling effective rapid prototyping and experimental validation. The key contributions of this work are:

- The implementation of low-cost and flexible SFCW radar architecture using SDR technology.
- Experimental validation of target detection accuracy in controlled indoor measurements.
- Analysis of system limitations to guide future improvements, including multipath effects and sidelobe interference.

This research demonstrates that SFCW-based SDR can provide a reliable, reconfigurable platform for radar research, and suitable for rapid prototyping.

https://doi.org/10.31436/iiumej.v27i2.3988

2. SYSTEM DESIGN

2.1. SFCW Signal Characteristics

SFCW radar operates by transmitting a train of sinusoidal signals that step uniformly in frequency. When the signals hit a target, they will bounce back to the radar, carrying the phase that determines the target's range. Fig. 2 and Fig. 3 show the key parameters of the SFCW signal. The signal consists of frequency components $(f_0, f_1, \ldots, f_{N-1})$ and these frequency components are separated by a fixed frequency step, Δf.

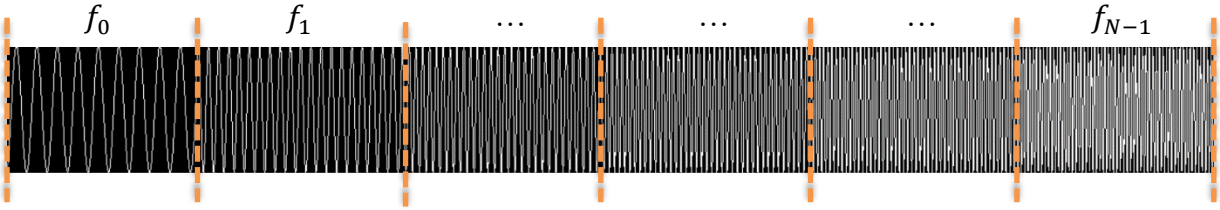


Figure 2. Illustration of the transmitted signals in SFCW radar

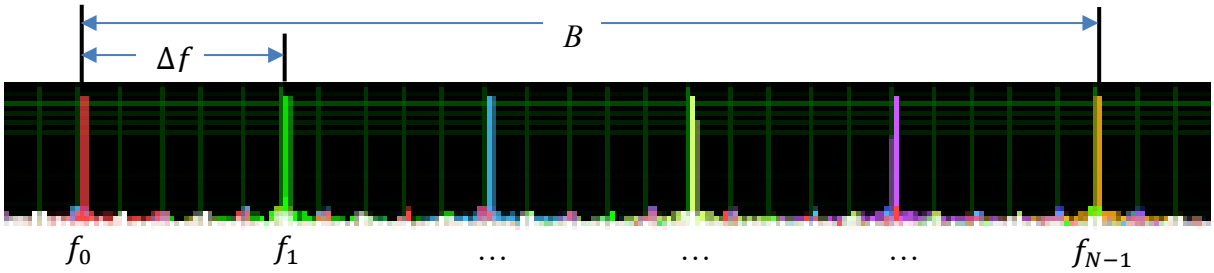


Figure 3. Frequency spectrum of the SFCW signal, highlighting the relationship between step size, number of frequency steps, and total bandwidth

The collection of stepped frequencies in SFCW spans a total bandwidth defined as

$$B = N\Delta f \tag{1}$$

where N is the number of frequency steps. The bandwidth is a critical parameter in an SFCW radar system, as it directly determines the system's range resolution, ΔR. Range resolution is the radar's ability to distinguish two closely spaced targets in range. The range resolution in SFCW radar is given by

$$\Delta R = \frac{c}{2B} \tag{2}$$

indicating that a larger bandwidth improves the system's ability to resolve targets that are close together.

Conversely, the maximum unambiguous range that can be detected using the SFCW system depends on the frequency step size, Δf, which is expressed as

$$R_{max} = \frac{c}{2\Delta f} \tag{3}$$

The SFCW approach is well-suited for far-field applications because its small frequency step size increases the unambiguous range, facilitating high-range target detection. However, this comes at the cost of requiring more frequency steps to maintain the same bandwidth. In this case, it would lead to increases in acquisition time and computational load. Thus, in SFCW

radar design, there is a trade-off between achieving high range resolution and maintaining sufficient unambiguous range.

In this research, the finalized design parameters for the SFCW radar system are summarized in Table 1. These values were selected to balance range resolution and unambiguous range while keeping the acquisition time practical for laboratory testing

Table 1. Design parameters of the SFCW radar system

Parameters	System Type	Length [m]
Start Frequency (f_0)	1.12 GHz	Initial transmit frequency
Stop Frequency (f_{N-1})	1.68 GHz	Final transmit frequency (f_0+ (N-1) Δf)
Frequency Step Size (Δf)	20 MHz	Step size between each frequency
Number of Steps (N)	28	Total number of frequency components
Total Bandwidth	560 MHz	$B = N\Delta f$
Range Resolution	ΔR	0.27 m
Unambiguous Range	R_{max}	7.5 m

With the system parameters defined, the next step is to show how to estimate the target range from the received signal.

The target range, R, can be estimated from the phase response, $\emptyset_N$ of the received signal at the frequency f_N, as shown in Eq. (4)

$$\emptyset_N = \frac{4\pi f_N R}{c} \tag{4}$$

From Eq. (4), the target range can be estimated directly from the phase of the received signal at each frequency component. However, in practice, the range profile is more conveniently obtained by applying an inverse fast Fourier transform (IFFT) to the entire frequency-domain dataset, thereby incorporating all frequency steps. In this research, the IFFT method was employed, with the peak locations in the range profile corresponding to the target distances.

To translate the design into practice, the SFCW radar system was implemented on an SDR platform as described in the following section.

2.2. Hardware Implementation

The proposed system employed a USRP N210 with a WBX daughterboard and dedicated transmit and receive antennas; the detailed hardware specifications are listed in Table 2.

Table 2. Design parameters of the SFCW radar system

Components	Details / Specifications
SDR Platform	USRP N210
Daughterboard	WBX (50 MHz – 2.2 GHz)
Max Sample Rate	25 MS/s
Transmit Antenna	Waveguide Antenna
Receive Antenna	Horn Antenna

To manage signal flow and hardware control, the system employed a dedicated software architecture developed in LabVIEW, as outlined in the next section.

https://doi.org/10.31436/iiumej.v27i2.3988

2.3. Software Architecture

The radar system was developed in LabVIEW, where all processes, including signal generation, USRP configuration, data acquisition, and initial signal processing, were carried out. LabVIEW was selected for this research because it offers seamless integration with the USRP hardware, graphical programming flexibility for waveform generation and hardware control, and built-in modules for real-time data acquisition. These features make it particularly suitable for rapid prototyping of SDR systems.

Fig. 4 shows the front panel of the developed user interface. The interface allows users to freely configure radar parameters, including operating frequency range, step size, number of frequency steps, and antenna selection. The flexibility of this interface is one of the main advantages of the developed system, as it can be rapidly adapted to experimental scenarios requiring varying radar configurations.

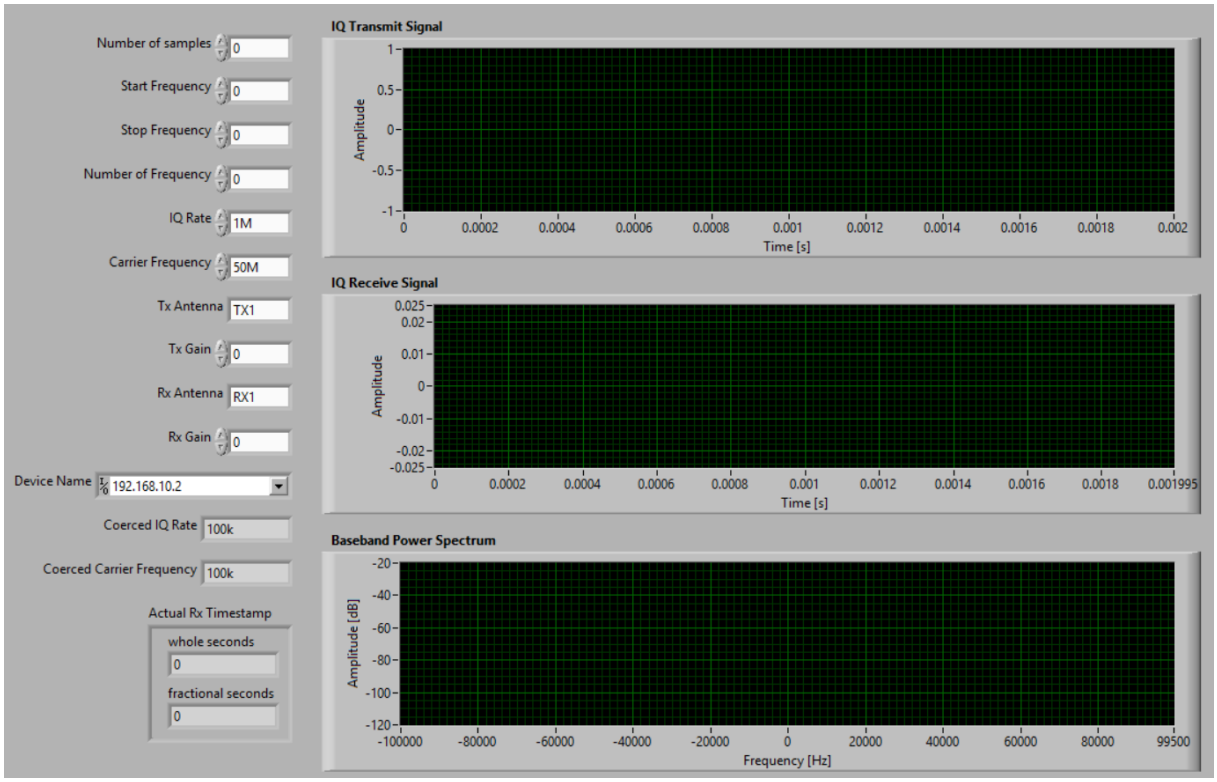


Figure 4. Example of LabVIEW user interface

The corresponding block diagram is shown in Fig. 5 and illustrates the graphical source code that defines the data acquisition process. The program is organized into three main sections. The first section covers waveform generation, the second covers the transmitter chain, and the third covers the receiver chain. This architecture ensures synchronized operation between the transmitter and receiver while maintaining flexibility for future modifications. In addition, the graphical nature of LabVIEW provides a clear visualization of the data flow, simplifying development, facilitating debugging, and enabling real-time monitoring, making it easier to identify and resolve potential system issues.

In summary, the proposed SFCW radar system was designed to balance range resolution and unambiguous range, implemented using USRP-based hardware, and controlled through a LabVIEW software architecture that ensured flexibility and ease of debugging. With the system design established, the following chapter describes the experimental methodology used to validate its performance.

https://doi.org/10.31436/iiumej.v27i2.3988

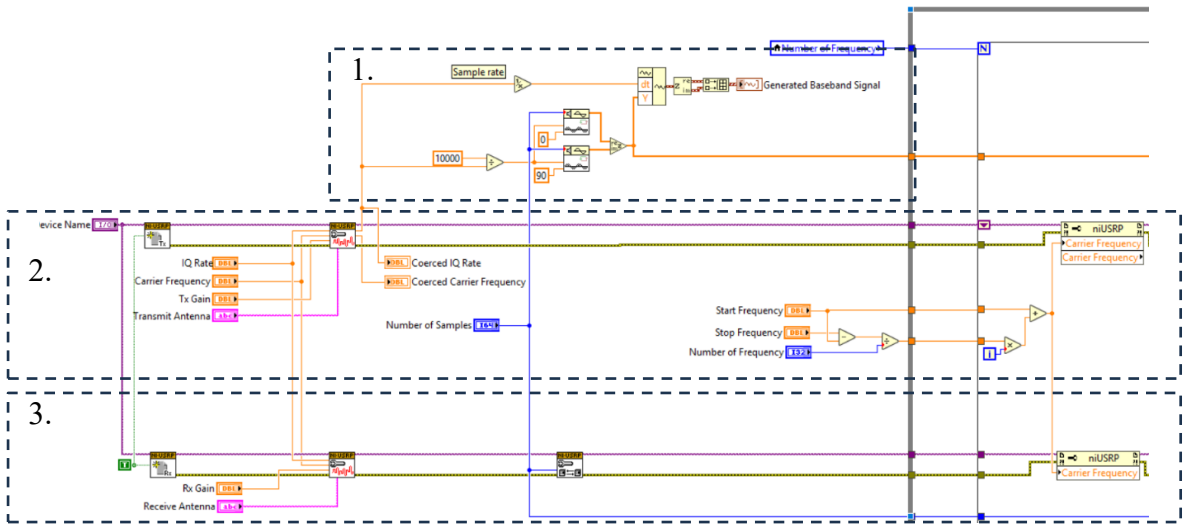


Figure 5. LabVIEW block diagram for the data acquisition process

3. METHODOLOGY

The system was tested to detect a 30 cm × 30 cm metal sheet in a controlled indoor environment, with the SFCW signal parameters summarized in Table 2. An overview of the radar system's transition from initialization to result visualization is shown in the flowchart in Fig. 6. The process comprises four stages: system setup, data acquisition, signal processing, and output.

As shown in Fig. 6, the procedure begins with system initialization, during which LabVIEW prepares the control interface and hardware connections. During the system setup stage, the USRP N210 is configured, and the operating parameters are defined as in Table 2.

The following system setup is the data-acquisition stage, in which SFCW signals are generated in LabVIEW and transmitted via the waveguide antenna. The reflected signals from the metal target are received by the horn antenna, downconverted to baseband IQ samples, and then stored as frequency-domain data for the signal processing stage.

Finally, in the Signal Processing & Output stage, the stored frequency-domain data is exported to LabVIEW for post-processing to obtain the range profile. The range profile enables estimation of the target position, which can then be compared with the actual distance. To obtain the range profile, an inverse fast Fourier transform (IFFT) is applied to the received signals, and the resulting spectrum is then displayed graphically.

To assess the accuracy of this estimation, the experimental setup and validation procedures are presented in the following subsections.

https://doi.org/10.31436/iiumej.v27i2.3988

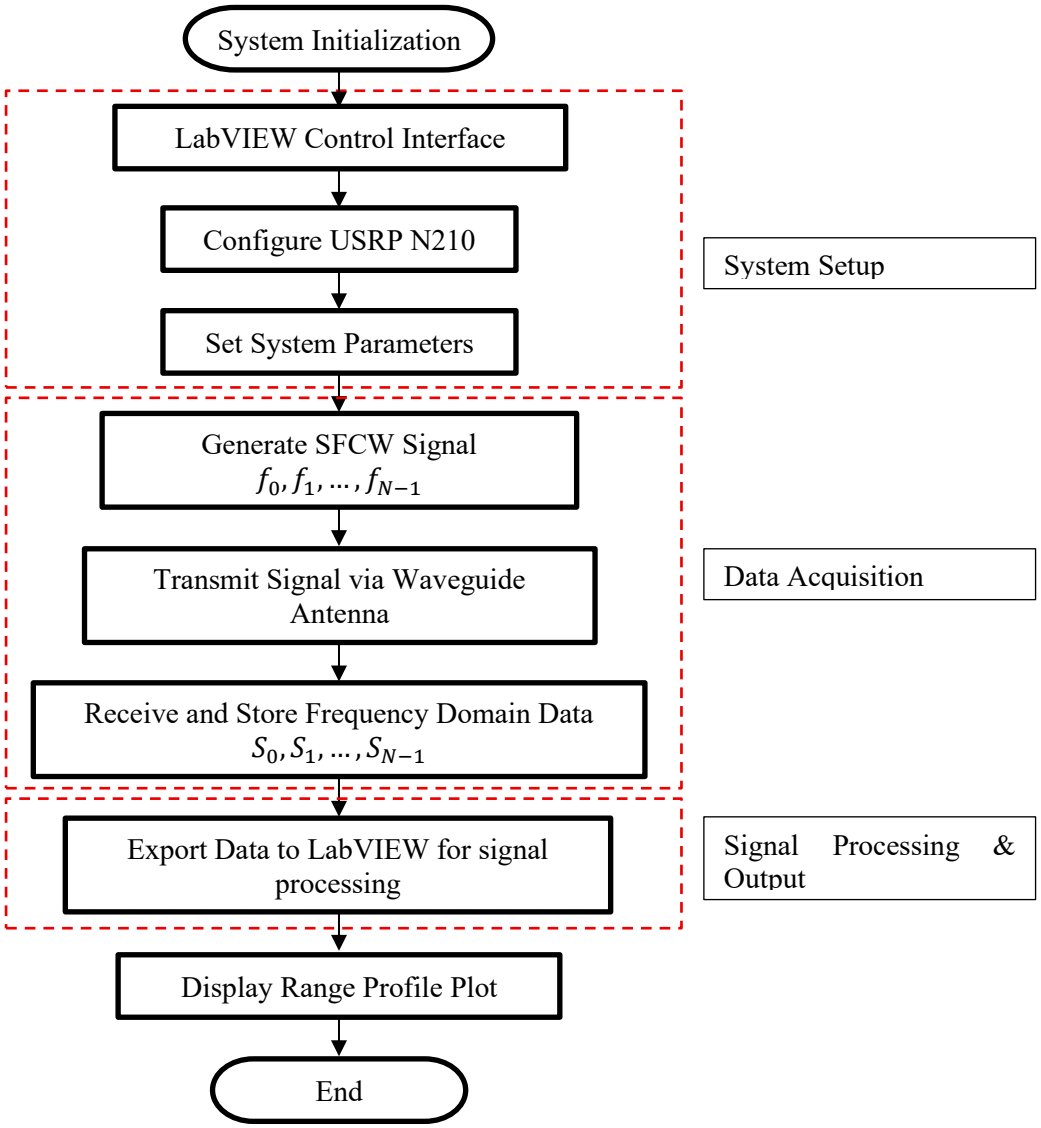


Figure 6. SFCW Radar System Processing Flow for Target Range Detection

3.1. System Setup

The proposed SFCW radar system was tested in a controlled indoor environment to minimize external interference. A metal sheet was used as the target and placed at various distances within the radar's detectable range. The system consisted of a USRP N210 with a WBX daughterboard, a waveguide antenna for transmission, and a horn antenna for reception, all connected to a host PC running LabVIEW for control and data acquisition.

Fig. 7 shows the experimental setup, with the transmitter and receiver antennas aligned toward the metal sheet target. Four tests were carried out, placing the target at 1.25 m, 1.00 m, 0.90 m, and 0.70 m, respectively. These distances were chosen to assess the radar's performance across a range of near-field and mid-range scenarios, enabling evaluation of its ability to accurately detect stationary objects at varying distances.

The experimental procedure began with initializing the radar system through the LabVIEW interface. The USRP N210 was first configured with the appropriate device address, sample rate, and carrier frequency, while transmit and receive antenna ports were assigned accordingly. The operating parameters summarized in Table 2 were set accordingly.

https://doi.org/10.31436/iiumej.v27i2.3988

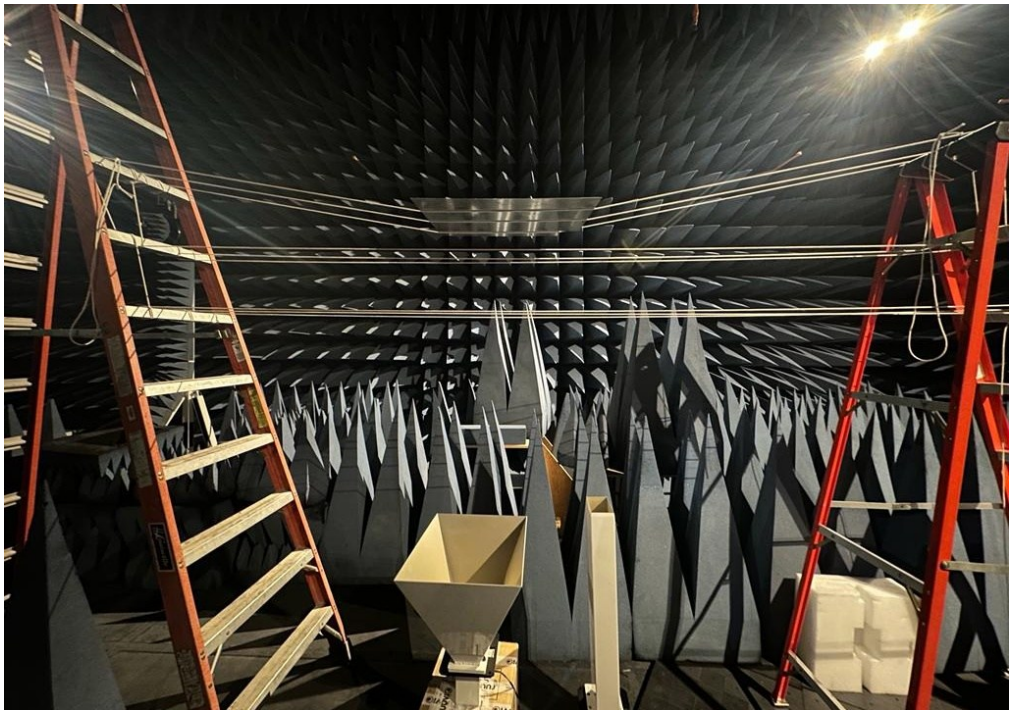
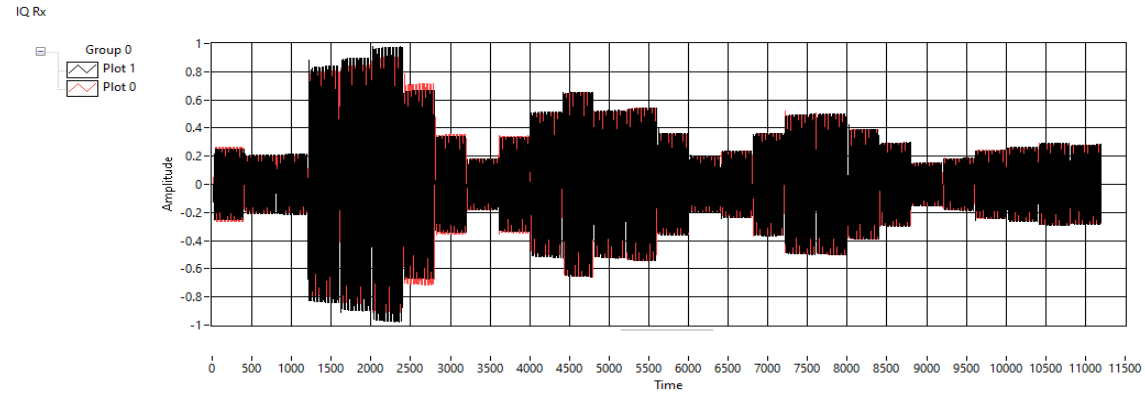


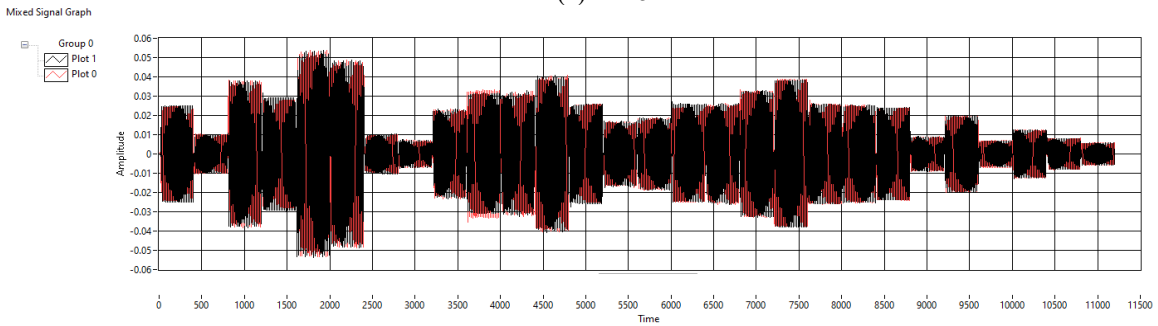
Figure 7. Experimental setup

3.2. Data Acquisition

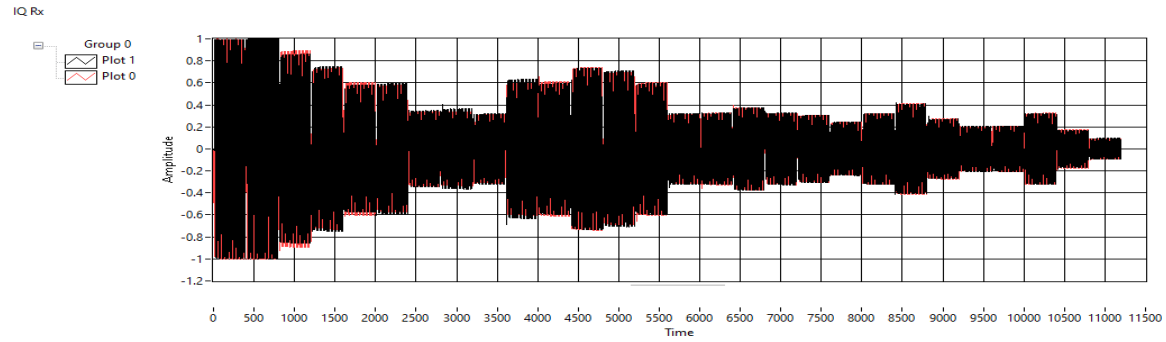
After system initialization, the radar generated SFCW signals, with each frequency step synthesized in LabVIEW and transmitted sequentially through the waveguide antenna. The reflected signals from the metal target were captured by the horn antenna and downconverted by the USRP N210 into baseband IQ samples.



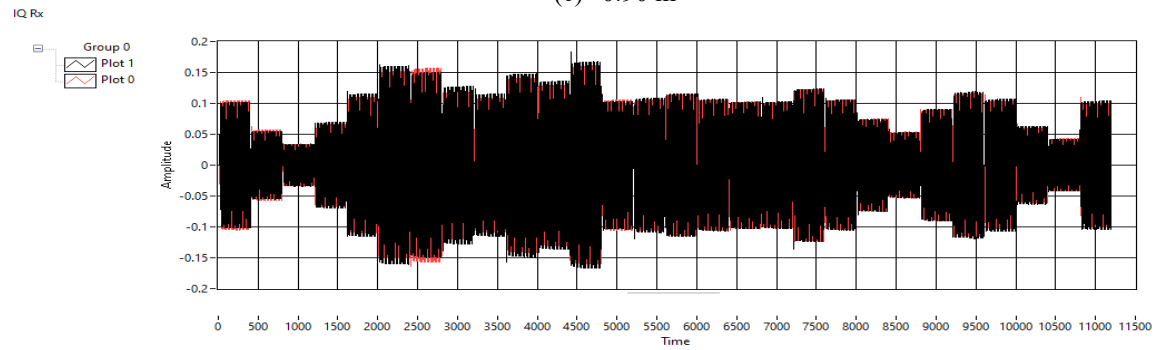
(a) 1.25 m



(b) 1.00 m



(c) 0.90 m



(d) 0.70 m

Figure 8. Received IQ data showing 28 frequency components for the target at (a) 1.25 m, (b) 1.00 m, (c) 0.90 m, and (d) 0.70 m

Fig. 8 (a-d) presents the received IQ data for the four measurement scenarios. In each case, 28 distinct frequency components are clearly visible, confirming that the system successfully received all transmitted frequencies. This validates the proper operation of the data acquisition process and ensures that the subsequent range-profile estimation is based on a complete set of frequency-domain measurements. The acquired data was then stored on the host PC for subsequent signal processing.

3.3. Signal Processing and Output

The I and Q samples received were saved in text format during acquisition and subsequently recalled within LabVIEW for post-processing. An inverse fast Fourier transform (IFFT) was applied across all 28 frequency steps to generate the target's range profile. The resulting range profile represents reflection power as a function of distance, with peaks corresponding to target locations.

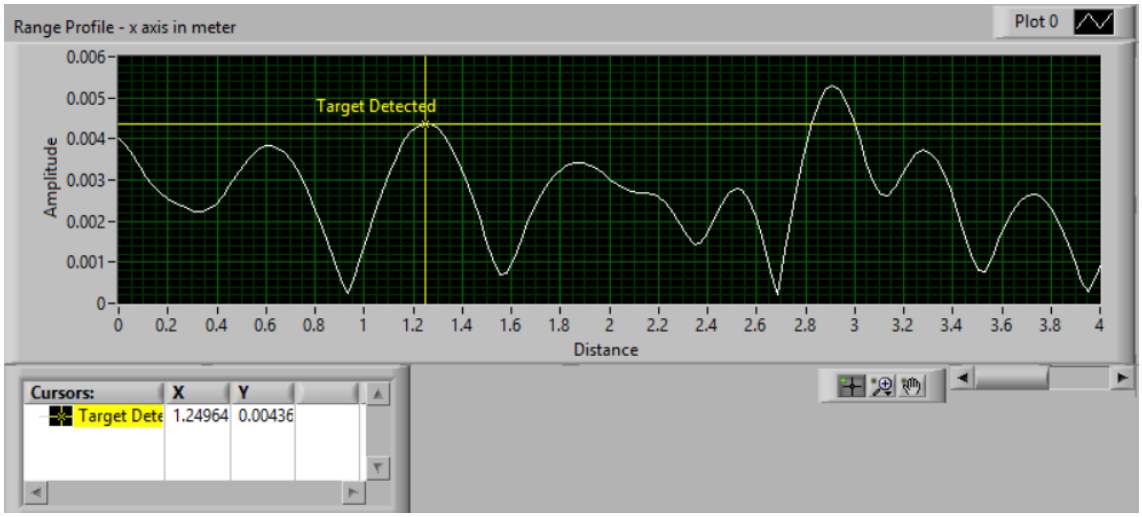
By using the full set of frequency steps, the IFFT-based method provides more accurate and robust range estimation than single-frequency phase analysis. This processing procedure served as the basis for evaluating the radar's performance, with representative range profiles presented in Section 4.

4. RESULTS

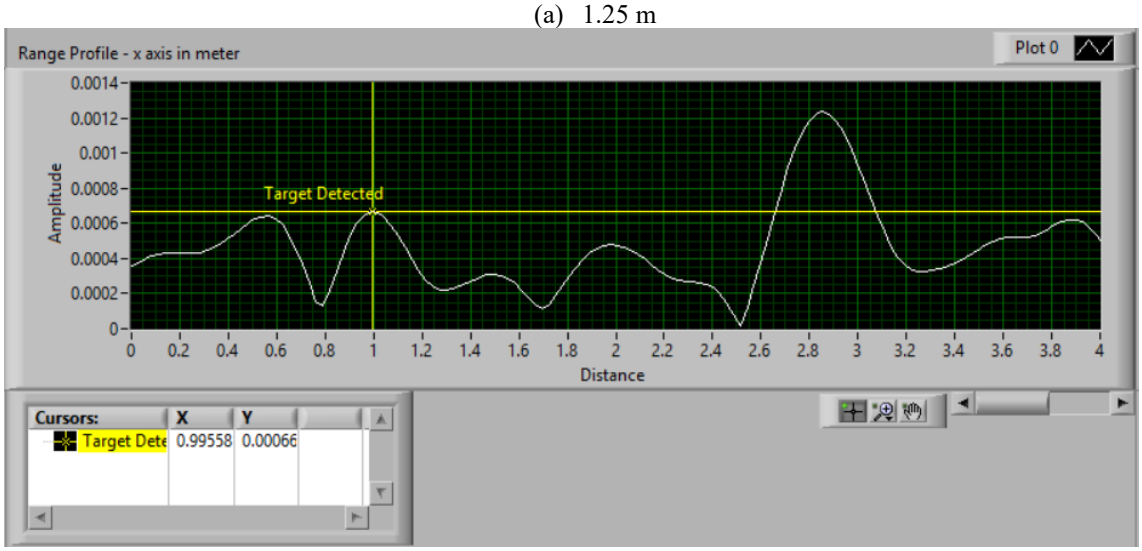
The radar system was evaluated by detecting a 30 × 30 cm metal target placed at four different ranges: 1.25 m, 1.00 m, 0.90 m, and 0.70 m. Fig. 9(a-d) shows the corresponding range profiles obtained via IFFT processing. The measured distances were compared with the actual distances, and the percentage of error was summarized in Table 3. Apart from the

https://doi.org/10.31436/iiumej.v27i2.3988

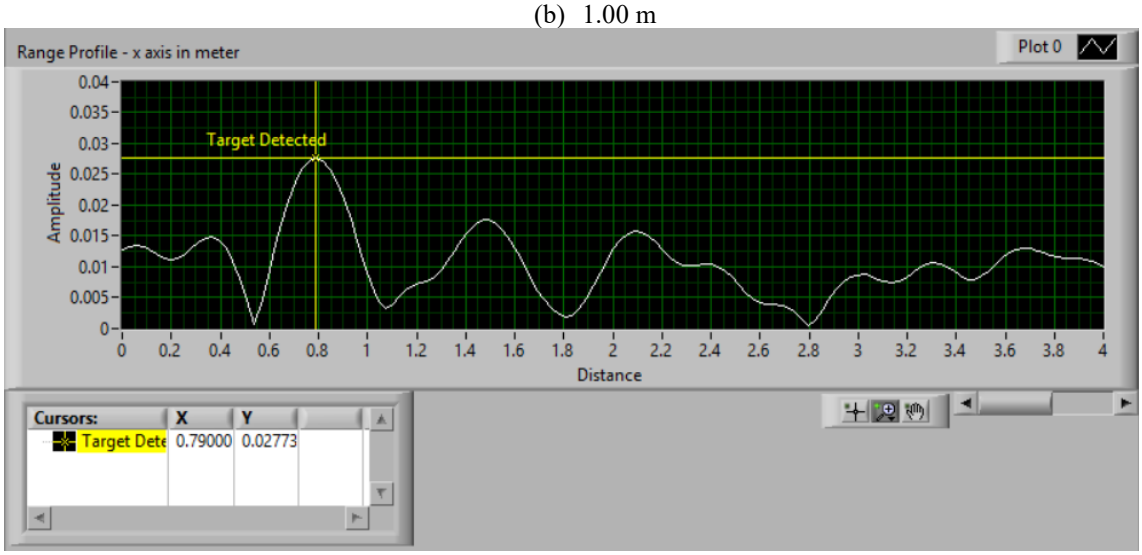
distance error, spurious spikes that appeared in the range profile also highlighted the need for further refinement of the system.



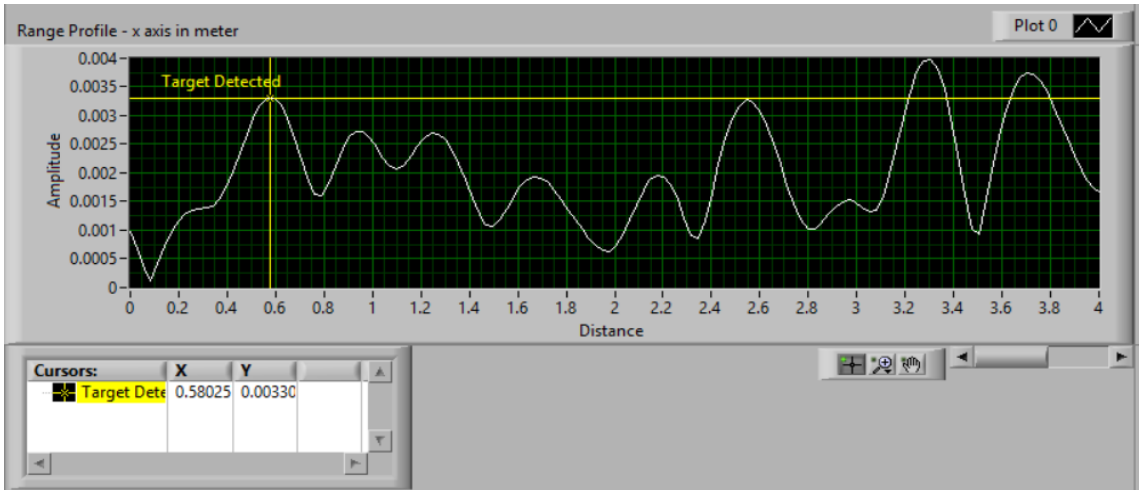
(a) 1.25 m



(b) 1.00 m



(c) 0.90 m



(d) 0.70 m

Figure 9. Range profile of the target placed at (a) 1.25 m, (b) 1.00 m, (c) 0.90 m, and (d) 0.70 m

Table 3. Design parameters of the SFCW radar system

Distance, m, (Actual)	Distance, m, (Measured)	%Error
1.25 m	1.24964 m ≈ 1.25 m	0.03% ≈ 0%
1.00 m	0.99558 m ≈ 1.00 m	0.44% ≈ 0%
0.90 m	0.79000 m ≈ 0.79 m	≈12.22%
0.70 m	0.58025 m ≈ 0.58 m	≈ 17.12%

The radar system demonstrated high accuracy at longer ranges of 1.25 m and 1.00 m, with approximately zero error recorded. However, at shorter ranges (0.9 m and 0.7 m), the percentage error increased to 12.22% and 17.12%, respectively.

5. DISCUSSION

The results demonstrate that the proposed SFCW SDR system can achieve accurate range estimation under controlled indoor conditions and reliably detect targets at 1.25 m and 1.00 m, highlighting the effectiveness of integrating the USRP N210 platform with LabVIEW for waveform generation, signal acquisition, and post-processing. It has successfully recorded errors of 0.03% and 0.44%, which are negligible.

The system's performance, however, degraded to shorter ranges of 0.90 m and 0.70 m, with percentage errors of 12.22% and 17.12%, respectively. Several factors may have contributed to these inaccuracies. First, in the experimental setup shown in Fig. 7, the ladder's position might cause multipath reflections in the laboratory that introduce additional echoes, which interfere with the main target response and shift peak positions in the range profile. Second, sidelobes in the IFFT response may have masked the main peaks, generated spurious spikes, and reduced detection clarity. Finally, imperfect calibration between the transmit and receive chains, along with residual system noise, may have affected phase alignment across frequency steps.

Based on the range profile observation in Fig. 9, the presence of spurious spikes further underscores the importance of advanced signal processing techniques. Target isolation could be improved by minimizing sidelobe effects and suppressing false peaks using window functions, averaging, or clutter suppression techniques. Additionally, near-field measurement accuracy would likely improve by calibrating and testing in environments with less multipath

https://doi.org/10.31436/iiumej.v27i2.3988

interference. Compared with FMCW-based SDR systems reported in prior studies, the proposed SFCW system offers key advantages, as summarized in Table 4.

Table 4. Performance comparison between the proposed SFCW system and existing FMCW SDR systems

Metric	[2] (2024) FMCW	[6] (2026) FMCW	[4] (2023) FMCW	This Research SFCW
Platform	USRP B210	LimeSDR	USRP 2932	**USRP N210 + WBX**
Bandwidth	14 MHz	58 MHz	80 MHz	**560 MHz**
Range Resolution	10.7 m	2.58 m	Not specified	**0.27 m**
Best Range Error	Not Specified	2.15% (avg, >15m)	Not specified	**0.03% (at 1.25 m)**
Near-field Error	Not specified	10-15% (<8 m)	Not specified	**0.44% (at 1.00 m)**
Power Feedthrough	Present	Present	Present (noted as critical)	**Absent**
Sweep Consistency	Inconsistent at wide BM	Requires reference path	Dependent on MIMO sync	**Stable (28/28 steps recovered)**
Software Control	GNU Radio	GNU Radio	LabVIEW	**LabVIEW (unified)**

From Table 4, the proposed SFCW system achieved a substantially finer range resolution of 0.27 m compared to 10.7 m in [2] and 2.58 m in [6]. This improvement shows a direct dependence on the wider effective bandwidth of 560 MHz enabled by the stepped-frequency architecture.

Meanwhile, in terms of detection accuracy, research in [6] reported average relative errors of 2.15% to 4.28% after applying a matched filter and a direct reference signal in an FMCW system on the LimeSDR platform over target distances of 5.27 m to 28.45 m. On the other hand, the proposed SFCW system achieved errors of 0.03% and 0.44% at 1.25 m and 1.00 m, respectively, without a reference signal path. In addition, it confirms that the stepped-frequency approach inherently provides more stable phase coherence across measurements.

Finally, research in [4] explicitly identified power feedthrough as a critical drawback of FMCW, which was addressed by using MIMO synchronization cable between two USRPs. In contrast, the proposed SFCW system has eliminated this issue by transmitting and receiving discrete step-frequency tones rather than continuous signals. Additionally, the SFCW system proposed in this research has demonstrated stability, as evidenced by a 100% frequency acquisition rate, with none of the 28 transmitted frequencies missing. On the other hand, in FMCW systems, achieving stability even at the equivalent bandwidth is difficult due to sweep-rate inconsistencies, as identified in [3]. The summary of the comparison of the existing SDR-based radar system and the proposed radar system is presented in Table 5.

On top of all, this research acknowledged that the proposed SFCW system introduces its own trade-off: in near-field detection, errors were recorded when the target was located at 0.90 m and 0.70 m, warranting further investigation. Nevertheless, the results collectively demonstrated that SFCW-based SDR offers measurable advantages over FMCW approaches for controlled, general-purpose indoor detection over sub-meter to meter ranges in terms of resolution, acquisition stability, and implementation simplicity.

Table 5. Comparative Analysis of Existing SDR-Based Radar Systems and the Proposed Approach

Research	Technique	Platform	Bandwidth	Key Performance Metrics	Limitation Addressed	Contribution & Significance
[2] (2024)	FMCW	USRP B210	14 MHz	Max range estimates 0.63 - 10.7 m (model dependent)	Limited transmit power and sweep time tradeoffs	Evaluates power constraints and sweep-time impacts on SDR-based FMCW radar
[4] (2023)	FMCW (SAR)	USRP 2932	80 MHz	SAR focusing demonstrated and hardware prototype	Power feedthrough and sweep rate instability in wideband	Validated real-time SAR imaging using MIMO cables for strict synchronization
[6] (2026)	FMCW	LimeSDR	58 MHz	Avg. Error: 2.15% - 4.28%	Random protocol delays and sync complexity in unsynchronized R_x	Developed matched filters and magnitude-squaring to handle unsynchronized SDR setups
[7] (2017)	SFCW (GPR)	USRP X310	4500 MHz	8-fold speed increase via multi-tone expansion	Slow data acquisition time and narrow instantaneous bandwidth	Expanded effective bandwidth of standard SDR hardware for GPR applications
[8] (2018)	SFCW (GPR)	Custom modular RF components	Not specified	Error: 0.2% - 2.0%	Lack of experimental verification for modular and portable radar designs	Validated portable SFCW-GPR for buried object detection
[9] (2021)	SFCW	USRP X300	5850 MHz	Validated SWE and snow depth measurements	Diffusion effects and complex layering in multi-path environments	Extended SFCW to environmental monitoring for hydrological cycle sensing
[10] (2020)	SFCW	bladeRF	5000 MHz	Sensitivity: 66 dB DR: 74 dB	High cost and redundant functionality of traditional VNAs	Confirmed SDR viability for precision medical imaging (breast cancer detection)
This study (2026)	**SFCW**	**USRP N210 + WBX**	**560 MHz**	**Error: <0.5% Res: 0.27m**	**Overcomes feedthrough, sweep instability, and synchronization complexity**	**Delivers low-cost, general-purpose SFCW SDR with sub-meter resolution and <0.5% mid-range error, outperforming FMCW baselines by an order of magnitude.**

Overall, the results highlight the strengths of SFCW SDR systems in achieving flexible, cost-effective, and accurate range detection, while also identifying the limitations that must be addressed to extend their applicability to real-world scenarios.

6. CONCLUSION

This work presented the implementation of a low-cost SFCW radar system using a USRP N210 platform controlled through LabVIEW. The system was experimentally validated in a controlled indoor environment, where it successfully detected a metal target at different ranges. The results showed that the system achieved nearly zero error for targets placed at 1.25 m and 1.00 m, confirming the effectiveness of the architecture and signal processing approach.

However, accuracy decreased at shorter ranges, with errors of 12.22% and 17.12% observed at 0.90 m and 0.70 m, respectively. These inaccuracies were primarily attributed to multipath reflections, sidelobe effects, and calibration limitations. While these findings highlight the system's limitations in near-field detection, practical improvements, such as advanced signal processing techniques and more robust calibration methods, have also been identified.

Overall, the research demonstrates that SFCW-based SDR systems provide a flexible and reliable platform for radar research. By combining software-defined control with the SFCW technique, such systems overcome many of the limitations of FMCW approaches and offer strong potential for adaptation to applications in environmental monitoring, biomedical sensing, and security. The near-field errors observed in this work highlight a critical boundary for Indoor Fall Detection and Posture Monitoring; therefore, subsequent studies will investigate the scaling of these models to complex, multi-target environments. Future iterations of this platform will integrate Machine Learning (ML) classifiers directly into the LabVIEW/software environment. By training a convolutional neural network (CNN) on SFCW range profiles, the system could move beyond simple detection to automated target recognition (ATR) and material classification, improving accuracy.

ACKNOWLEDGEMENT

This research was supported by the International Islamic University Malaysia (IIUM) and the Ministry of Higher Education Malaysia (MOHE) through the Fundamental Research Grant Scheme (FRGS/1/2024/TK08/UIAM/02/1).

REFERENCES

[1] Costanzo, S., Spadafora, F., Di Massa, G., Borgia, A., Costanzo, A., Aloi, G., Pace, P., Loscrì, V., & Moreno, H. O. (2013). Potentialities of usrp-based software defined radar systems. *Progress In Electromagnetics Research B*, *53*, 417–435. https://doi.org/10.2528/PIERB13052904

[2] Putra, B. P. H., Akbar, F. S., & Untsa, R. M. (2024). Maximum Range Analysis of An SDR-Based FMCW Radar. 2024 IEEE 22nd Student Conference on Research and Development, SCOReD 2024, 309–312. https://doi.org/10.1109/SCOReD64708.2024.10872711

[3] Nguyen, C., & Park, J. (2016). Stepped-Frequency Radar Sensors Theory, Analysis and Design. SpringerBriefs in Electrical and Computer Engineering. https://doi.org/10.1007/978-3-319-12271-7_3

[4] Ashry, M. M., & Mashaly, A. S. (2023). FMCW SAR Imaging Based on Software Defined Radio Architecture. 2023 International Telecommunications Conference, ITC-Egypt 2023, 731–736. https://doi.org/10.1109/ITC-Egypt58155.2023.10206395

[5] Suksmono, A. B. (2013). A Simple Solution To The Uncertain Delay Problem in USRP Based SDR-Radar Systems. 10, 1–4.

[6] Juan Carlos, Q. M., & Paul Edward, P. G. (2026). Range estimation in FMCW SDR radars with unsynchronized receivers. Physical Communication, 75(October 2025), 103000. https://doi.org/10.1016/j.phycom.2026.103000

[7] Carey, S. C., & Waymond R. Scott, J. (2017). SOFTWARE DEFINED RADIO FOR STEPPED-FREQUENCY, GROUND-PENETRATING RADAR. 4825–4828.

[8] Arifianto, M. S., Hakim, N. F. A., Chairunnisa, & Munir, A. (2019). Experimental Analysis of Range Detection for Portable SFCW Ground Penetrating Radar. Mediterranean Microwave Symposium, 2018-Octob, 372–374. https://doi.org/10.1109/MMS.2018.8612115

[9] Alonso, R., García del Pozo, J. M., Buisán, S. T., & Álvarez, J. A. (2021). SNOW WATER EQUIVALENT EVOLUTION DURING THE 2019 / 2020 WINTER PERIOD IN AEMET-FORMIGAL TEST SITE USING A SFCW RADAR. 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 1035–1038. https://doi.org/10.1109/IGARSS47720.2021.9554022

[10] Carvalho, D., Aragao, A. J., Ferrari, A., Sanches, B., & Noije, W. (2020). Software-Defined Radio Assessment for Microwave Imaging Breast Cancer Detection. 2020 IEEE Nordic Circuits and Systems Conference, NORCAS 2020 - Proceedings. https://doi.org/10.1109/NorCAS51424.2020.9265007

https://doi.org/10.31436/iiumej.v27i2.4008

Raindrop Size Distribution and Wind-Driven Rain Specific Attenuation: CCDF-Based Analysis at 18 and 38 GHz in Tropical Country

NURUL NAJWA MD YUSOF[1,3], OMAR ABDUL AZIZ[1]*, LAM HONG YIN[2], JAFRI DIN[1]

[1]*Faculty of Engineering, Universiti Teknologi Malaysia, Johor Bahru, Malaysia*
[2]*Faculty of Engineering Technology, Universiti Tun Hussein Onn Malaysia, Batu Pahat, Malaysia*
[3]*Polytechnic and Community College Education Department, Ministry of Higher Education, Putrajaya, Malaysia*

**Corresponding author: omar@utm.my*

(Received: 23 October 2025; Accepted: 8 March 2026; Published online: 10 May 2026)

ABSTRACT: Seasonal monsoons, frequency dependence, and wind effects were investigated for rain-specific attenuation at millimeter-wave (mm-wave) frequencies in a tropical environment. Using one-minute RD-69 DSD data and matched wind observations, CCDFs of specific attenuation were evaluated at 18 and 38 GHz for three monsoon phases (SW, NE, IM) and four wind-direction sectors under unconstrained and constrained winds (9–11 m/s). The CCDF results show clear seasonal variability, with the SW monsoon producing the strongest attenuation and the largest mismatch from ITU-R P.838-3, particularly at 38 GHz. Error statistics show that the maximum deviation occurs in the SW 270°–360° sector, with an RMSE of 1.550 dB/km and a positive bias, confirming ITU-R underestimation. Meanwhile, the NE monsoon exhibits smaller overall deviations (worst-case RMSE below about 0.6 dB/km at 38 GHz). The IM period exhibits mixed behavior, where wind direction can shift the deviation between under- and overestimation under constrained winds. Exceedance levels of 0.01% and 0.001% were emphasized as design-relevant extremes, corresponding to approximately 52.6 min/year and 5.26 min/year, respectively, under one-minute sampling. Overall, the results indicate that ITU-R P.838-3 accuracy in tropical conditions depends on monsoon phase and wind direction, supporting monsoon-aware, direction-sensitive fade-margin planning and link-orientation considerations for robust 5G/6G terrestrial mm-wave systems.

ABSTRAK: Monsun bermusim, kebergantungan frekuensi, dan kesan angin telah dikaji terhadap pelemahan spesifik hujan pada frekuensi gelombang milimeter (mm-wave) dalam persekitaran tropika. Menggunakan data DSD RD-69 beresolusi satu minit dan melalui pemerhatian padanan angin, pelemahan CCDF spesifik dinilai pada 18 dan 38 GHz untuk tiga fasa monsun (SW, NE, IM) dan empat sektor arah angin di bawah angin tidak terkekang serta angin terkekang (9–11 m/s). Penyimpangan terbesar daripada ITU-R P.838-3 berlaku pada 38 GHz, khususnya semasa monsun barat daya (SW), apabila sektor 270°–360° di bawah angin tidak terkekang menunjukkan RMSE = 1.550 dB/km dan bias = +1.193 dB/km, menandakan ITU-R bernilai lebih rendah; di bawah angin terkekang, penyimpangan kekal tinggi (RMSE = 1.383 dB/km). Tempoh peralihan (IM) juga menunjukkan kebergantungan pada sektor angin terkekang, dengan RMSE = 1.212 dB/km dan bias = +0.936 dB/km pada 38 GHz, manakala monsun timur laut (NE) menunjukkan penyimpangan lebih kecil (RMSE terburuk adalah kurang daripada kira-kira 0.6 dB/km pada 38 GHz di bawah angin tidak terkekang). Tahap kebarangkalian lebihan 0.01% dan 0.001% ditekankan sebagai ekstrem yang relevan bagi reka bentuk, bersamaan kira-kira 52.6 min/tahun dan 5.26 min/tahun, masing-masing, pada pensampelan satu minit. Secara keseluruhan, dapatan kajian menunjukkan ketepatan ITU-R P.838-3 dalam keadaan tropika bergantung pada fasa monsun dan arah angin,

sekaligus menyokong perancangan margin-pudaran yang peka monsun dan peka arah angin serta menimbang orientasi pautan untuk sistem gelombang-mm 5G/6G terestrial yang lebih berdaya tahan.

KEYWORDS: *raindrop size distribution, DSD, specific attenuation, millimeter-wave*

1. INTRODUCTION

Millimeter-wave (mm-wave) link development is gaining attention for its application in fifth- and sixth-generation (5G/6G) high-capacity cellular services. Nevertheless, communication in this frequency range is susceptible to rain attenuation, particularly in tropical environments with convective thunderstorms that produce very large raindrops. Absorption and scattering of wireless signals are affected by raindrop size as well as the signal's wavelength. Given that specific attenuation increases with frequency, the 18–40 GHz spectrum is particularly vulnerable to the effects of tropical rain microphysics [1]. Communication links operating at 38 GHz and beyond are further degraded by depolarization caused by oblate convective raindrop shapes during thunderstorms [2]. The aforementioned scenarios are typical of Malaysia's tropical climates. Considering the standard 1-minute time interval, Malaysia experiences rain rates exceeding 100 mm/h for 0.01% of the time, which is among the highest in the world [3-4]. Local Malaysian measurements recorded a greater concentration and wider range in the raindrop size distribution (DSD) than in temperate climates. This contributes to the rainy season's decaying profile with higher specific attenuation at 18/38 GHz [5-6], As reported in [7-8], rain-induced attenuation measurements at 18 GHz and 38 GHz recorded 33 dB and 21 dB, respectively, during the 0.01% worst-case operational time in Malaysia. Since most standard prediction tools for mm-wave links, including models from the International Telecommunication Union Radiocommunication (ITU-R), are primarily based on temperate-region data, calibration and correction of these models with respect to tropical data are necessary [9-10].

The 38 GHz band has been allocated for mm-wave 5G [7], while the 18 GHz band is under serious consideration for the upcoming 6G [11]. Verifying the channels' potential coverage distance is therefore critical when planning future cellular networks [12-13]. Given the importance of both bands for terrestrial point-to-point and fixed network access systems, their operation has been considered in the ITU-R P.530 standard model when rainfall occurs [14, 14]. However, for prediction models to closely resemble actual cases, rain rate and DSD data for the specific region should be used based on tropical climate samples [16-17]. Most previous works have concentrated on frequency scaling and variability driven by DSD and have largely treated precipitation as isotropic and time-invariant, which is not the case. Wind, in both speed and direction, as well as the monsoon phase, dramatically influences the shape, orientation, and alignment of raindrops. This creates direction-dependent attenuation that most models do not account for [18-20]. These gaps can cause systematic fade-margin shortfalls during extreme rain events, putting link reliability at risk of not meeting the availability targets expected of 5G and 6G backhaul.

The need to investigate the combined influence of the monsoon season, wind direction, and wind speed on specific attenuation at 18 GHz and 38 GHz is therefore essential. This paper closes this gap by investigating rain attenuation, accounting for seasonal monsoon, wind vectors, and DSD microphysics to quantify exceedance-level behavior. This study aims to provide a comprehensive assessment of standard rain-specific attenuation in tropical conditions, namely the ITU-R model, to formulate a monsoon-aware, direction-dependent fade-margin framework that improves 5G link design and future sub-terahertz 6G system

planning [6,7,21]. The remainder of this paper is organized as follows. Section 2 presents the methodology, including the RD-69 DSD dataset, the data extraction strategy, wind-constraint classification, the specific attenuation computation using both ITU-R P.838-3 and the DSD-derived specific attenuation, and the CCDF construction procedure. Section 3 reports and discusses the CCDF results across monsoon phases and wind-direction sectors, supported by RMSE and bias statistics for model comparison. Section 4 concludes the paper by summarizing the key findings and their implications for mm-wave fade-margin planning and link-orientation considerations for 5G/6G deployments in tropical climates.

2. METHODOLOGY

Three years of DSD data in this study were measured using the RD-69 Joss-Waldvogel disdrometer model. This disdrometer is located at Universiti Teknologi Malaysia, Kuala Lumpur, and is capable of measuring rainfall attributes, drop size, and the interval to enable high-resolution measurements. It identifies and classifies individual drops by size, from 0.36 to 5.30 mm, and by fall velocity as they traverse a sensing zone, sorting them into 20 different bins. The RD-69 disdrometer captures finer details of the DSD over time and ensures reliable, precise measurements. For the RD-69 disdrometer, a conical styrofoam collector is positioned over the sensing unit. This collector reduces wind effects and ensures that raindrops fall vertically onto the sensor. The sensing unit is a coil of wire suspended within a magnetic field.

A moving coil is the result of a droplet striking the collector. This movement activates the sensor, and a designed measuring device is created, where the data logger captures the corresponding voltage. The disdrometer captures the DSD from the voltage pulses produced by the RD-69 disdrometer. The size of each pulse indicates the size of the corresponding raindrop. The number of pulses corresponding to each size category is divided by the sampling interval to obtain the rainfall rate in mm/h. More information on the measurement setup is described in [22]. To ensure consistency across seasons, the dataset was grouped into three monsoon phases: Southwest (SW) monsoon, Northeast (NE) monsoon, and Intermonsoon (IM). Wind direction was further grouped into four sectors (0°–90°, 90°–180°, 180°–270°, and 270°–360°) to enable a directional comparison of rain microphysics and the corresponding specific attenuation.

2.1. Data Extraction Strategy

For each monsoon phase and wind direction, contiguous 20-minute windows were extracted from the one-minute DSD and wind records. This window length is suitable for tropical rainfall because it is short enough to capture rapid temporal changes in rainfall intensity and the associated evolution of the DSD during different phases of a rain event, yet long enough to provide a sufficient sample size for stable exceedance statistics after stratification [23]. To ensure comparability across categories and to focus on operationally relevant intense rain, only windows with a mean rainfall rate of 40–50 mm/h were retained, where the mean was computed as the arithmetic average of the 20 one-minute rainfall rates. Limiting the analysis to a narrow rain-rate range also makes the comparison fairer because the results are less influenced by changes in rainfall intensity and more related to DSD variability and wind conditions [24]. Windows with missing DSD or wind records were discarded. Wind speed and wind direction were obtained from the Malaysian Meteorological Department (MET Malaysia) and matched to the same observation periods as the DSD records. Tables 1–3 summarize the average drop counts for the SW, NE, and IM periods, respectively, stratified by wind-direction sector under both unconstrained winds and constrained winds (9–11 m/s).

https://doi.org/10.31436/iiumej.v27i2.4008

Table 1. Average drop counts during the SW monsoon under unconstrained winds and constrained winds (9–11 m/s).

Bin Number	Mean Diameter [mm]	Unconstrained winds				Constrained winds (9-11 m/s)			
		SW				SW			
		0° - 90°	90° - 180°	180° - 270°	270° - 360°	0° - 90°	90° - 180°	180° - 270°	270° - 360°
1	0.36	7	4	6	4	7	1	7	0
2	0.46	10	17	13	14	10	1	12	1
3	0.55	17	28	17	35	17	4	19	8
4	0.66	41	44	26	64	41	5	35	23
5	0.77	50	45	31	75	50	7	43	33
6	0.91	118	105	90	139	118	43	118	95
7	1.12	183	175	168	189	183	117	188	141
8	1.33	142	137	123	142	142	99	144	108
9	1.51	105	97	90	107	105	71	100	78
10	1.67	97	88	79	108	97	60	80	65
11	1.91	135	142	105	159	135	83	115	91
12	2.26	93	91	70	121	93	62	80	74
13	2.58	43	42	38	59	43	38	41	39
14	2.87	27	27	26	33	27	23	29	24
15	3.20	21	22	21	24	21	23	27	22
16	3.54	13	11	12	11	13	14	15	15
17	3.92	6	8	12	5	6	13	11	9
18	4.35	2	3	7	2	2	9	5	4
19	4.86	1	2	4	1	1	6	2	3
20	5.30	0	1	2	0	0	3	1	1

2.2. Wind Constraint Classification

To assess the effect of wind, the DSD samples were separated into two groups, considering wind speed: (i) unconstrained wind, which includes all retained segments regardless of wind speed, and (ii) constrained wind, which includes only segments with wind speeds of 9–11 m/s. Wind speeds in Malaysia are influenced by monsoon seasons, with significant variability throughout the year. The NE monsoon, in particular, brings higher wind speeds. Selecting a range of 9–11 m/s would capture the upper end of typical wind speeds during these periods, providing a realistic scenario for studying wind effects on mm-wave propagation. Furthermore, wind speeds in this range can cause antenna misalignment due to wind-induced vibrations, which is critical for maintaining the reliability of mm-wave communication links [25]. Therefore, analyzing this regime provides practical guidance for mitigation and more robust link planning.

2.3. Specific Attenuation Computation

This study assesses two methods to calculate specific attenuation. Specific attenuation, γ (dB/km), was computed using two approaches: (i) the standard ITU-R P.838-3 power-law model, and (ii) a DSD-derived specific attenuation based on the forward scattering amplitude (FSA) obtained via the T-matrix method. The ITU-R standard has established a widely recognized method for calculating specific attenuation as a function of frequency and rain rate. This method offers a reliable means of estimating the attenuation of the rain-induced signal. The ITU-R P.838-3 equation can be written as described in Eq. (1), where R is the rainfall rate (mm/h), and k and α are frequency and polarization-dependent.

$$\gamma = kR^{\alpha} \text{ (dB/km)} \quad (1)$$

https://doi.org/10.31436/iiumej.v27i2.4008

Specific attenuation was also computed directly from the measured DSD using a T-matrix approach. In this method, the complex forward-scattering amplitude (FSA) is computed for each drop-diameter bin and summed over all bins to obtain the specific attenuation per unit length, as given in Eq. (2). This DSD-derived specific attenuation captures microphysical variability that is not represented by rainfall-rate-only models and enables a physically consistent comparison with the ITU-R prediction.

$$\gamma_{H,V} = 8.686 \times 10^3 \times \lambda \times \sum_{i=1}^{20} Im\{f_{H,V}(D_i)\} \cdot N(D_i) \cdot \Delta D_i \text{ (dB/km)} \quad (2)$$

where λ is the wavelength (m), $Im\{\cdot\}$ denotes the imaginary part, $f_{H,V}(D_i)$ is the T-matrix derived forward-scattering amplitude for drop diameter D_i, $N(D_i)$ is the drop concentration for bin i ($m^{-3}mm^{-1}$), and ΔD_i is the bin width in mm. The constant in (2) provides conversion to dB/km under the units defined above. The T-matrix method is used to compute $Im(f_{H,V}(D_i))$ for each diameter bin, allowing the attenuation to be driven directly by the observed DSD rather than by rainfall rate alone [6].

The bin-wise drop-count distributions used to support the DSD-derived attenuation calculation are summarized in Tables 2 and 3. Table 2 presents the average drop counts during the northeast (NE) monsoon, while Table 3 presents the corresponding values during the inter-monsoon (IM) period. In both tables, the drop counts are grouped according to mean drop diameter, wind-direction sector, and wind-speed condition, namely unconstrained winds and constrained winds within 9–11 m/s. These distributions provide the microphysical basis for estimating $N(D_i)$ in Eq. (2), enabling the attenuation calculation to reflect seasonal and wind-dependent variations in the observed DSD.

Table 2. Average drop counts during the NE monsoon under unconstrained and constrained winds (9–11 m/s).

Bin Number	Mean Diameter [mm]	Unconstrained winds				Constrained winds (9-11 m/s)			
		NE				NE			
		0° - 90°	90° - 180°	180° - 270°	270° - 360°	0° - 90°	90° - 180°	180° - 270°	270° - 360°
1	0.36	9	3	10	1	10	6	0	6
2	0.46	40	10	31	6	25	17	2	14
3	0.55	71	19	38	23	28	27	14	20
4	0.66	98	40	56	51	33	51	36	30
5	0.77	93	44	53	60	29	56	61	29
6	0.91	181	114	132	135	97	131	134	74
7	1.12	270	199	210	188	220	214	201	152
8	1.33	186	140	163	136	166	172	133	133
9	1.51	119	102	99	102	116	124	94	97
10	1.67	105	97	91	93	100	110	91	82
11	1.91	141	139	121	135	123	147	141	122
12	2.26	97	92	79	103	77	96	99	79
13	2.58	45	44	37	46	38	45	56	41
14	2.87	26	27	24	26	24	29	34	28
15	3.20	22	19	22	20	21	21	26	24
16	3.54	11	10	11	9	15	8	11	12
17	3.92	7	6	9	7	11	6	6	9
18	4.35	3	2	4	3	4	2	2	4
19	4.86	1	1	2	1	2	1	0	2
20	5.30	0	0	1	0	0	0	0	1

https://doi.org/10.31436/iiumej.v27i2.4008

Table 3. Average drop counts during IM monsoon under unconstrained winds and constrained winds (9–11 m/s).

Bin Number	Mean Diameter [mm]	Unconstrained winds				Constrained winds (9-11 m/s)			
		IM				IM			
		0° - 90°	90° - 180°	180° - 270°	270° - 360°	0° - 90°	90° - 180°	180° - 270°	270° - 360°
1	0.36	6	20	1	0	4	3	5	1
2	0.46	24	38	8	4	12	11	14	4
3	0.55	36	39	48	19	21	26	21	18
4	0.66	62	58	90	47	40	45	32	43
5	0.77	59	62	94	52	50	47	38	51
6	0.91	126	154	184	118	120	106	95	117
7	1.12	192	247	230	182	189	184	164	186
8	1.33	141	177	170	134	153	141	136	131
9	1.51	97	113	109	99	116	97	96	100
10	1.67	87	94	97	93	112	86	86	95
11	1.91	128	131	131	141	180	120	130	144
12	2.26	89	92	88	93	115	78	95	95
13	2.58	45	43	48	40	56	42	43	42
14	2.87	32	25	29	27	32	30	23	28
15	3.20	27	23	23	19	18	26	20	21
16	3.54	13	12	9	8	6	12	11	10
17	3.92	7	7	5	6	3	8	8	7
18	4.35	2	3	1	3	1	3	4	4
19	4.86	0	1	0	2	0	1	2	2
20	5.30	0	0	0	1	0	0	1	1

2.4. CCDF Construction

In this study, the complementary cumulative distribution function (CCDF) was used to quantify the frequency with which the specific attenuation exceeds a given level under different monsoon and wind conditions. First, one-minute DSD measurements from the RD-69 disdrometer were used to derive a time series of specific attenuation, $\gamma(t)$, for each retained 20-minute window. The specific attenuation was computed using Eq. (1) and Eq. (2). The resulting $\gamma(t)$ samples were then grouped according to the analysis categories (monsoon phase, wind direction, and wind speed group). For each category, all specific attenuation samples were pooled and sorted, and the exceedance probability for any specific attenuation threshold x was evaluated as $P(\gamma \geq x)$, where $P(\cdot)$ denotes probability. In practice, it was estimated as the fraction of samples with γ greater than or equal to x. Plotting $P(\gamma \geq x)$ against x produced the CCDF curves, which were used to compare the DSD-derived specific attenuation statistics with the corresponding ITU-R predictions under the same conditions. Finally, exceedance levels at 0.01% and 0.001% were reported as reference points for link design. With one-minute sampling, these correspond to approximately 52.6 min/year and 5.26 min/year, respectively.

3. Results and Discussion

This study evaluates how monsoon phase and wind direction influence mm-wave specific attenuation in a tropical environment and whether ITU-R P.838-3 remains adequate when attenuation is derived from local DSD. The CCDF results show that agreement with ITU-R is not uniform across seasons or wind sectors, and the deviation is most pronounced at 38 GHz. This matters because 38 GHz is a key FR2 band where fade margins are typically tight; even sector-dependent differences of the order of 1 dB/km can accumulate into noticeable

https://doi.org/10.31436/iiumej.v27i2.4008

uncertainty for short-path terrestrial planning, especially at low time percentages relevant to high availability targets. To complement the CCDF comparisons, Tables 4 and 5 summarize the model deviations in terms of RMSE and bias between the DSD-derived attenuation and ITU-R P.838-3. A positive bias (+) indicates ITU-R underestimation, while a negative bias (−) indicates overestimation. In this context, bias explains the *direction* of the mismatch, whereas RMSE reflects the typical *magnitude* of the mismatch in dB/km across the evaluated sectors. This distinction is important because bias can appear small when under- and overestimation across directions partially cancel out, while RMSE still reveals substantial spread across sectors.

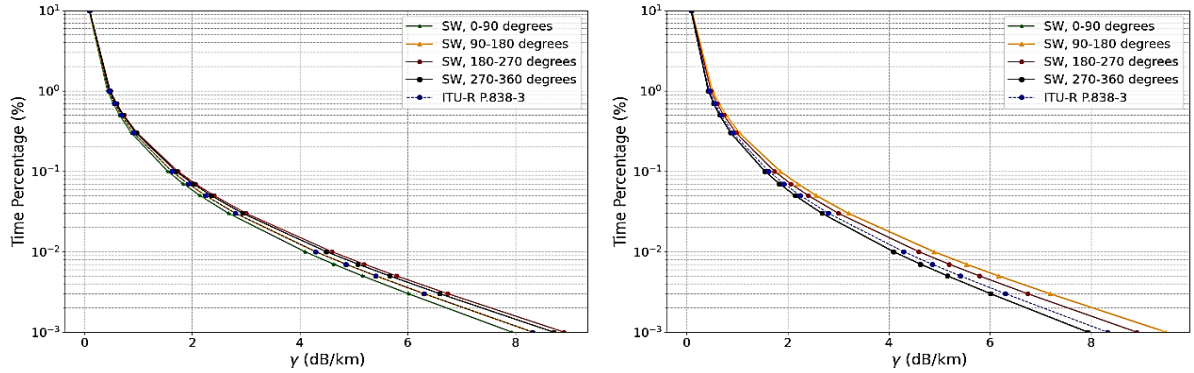


Figure 1. CCDF of specific attenuation (γ, dB/km) during the SW monsoon for 40–50 mm/h against ITU-R P.838-3 at 18 GHz under (a) unconstrained winds and (b) constrained winds (9–11 m/s)

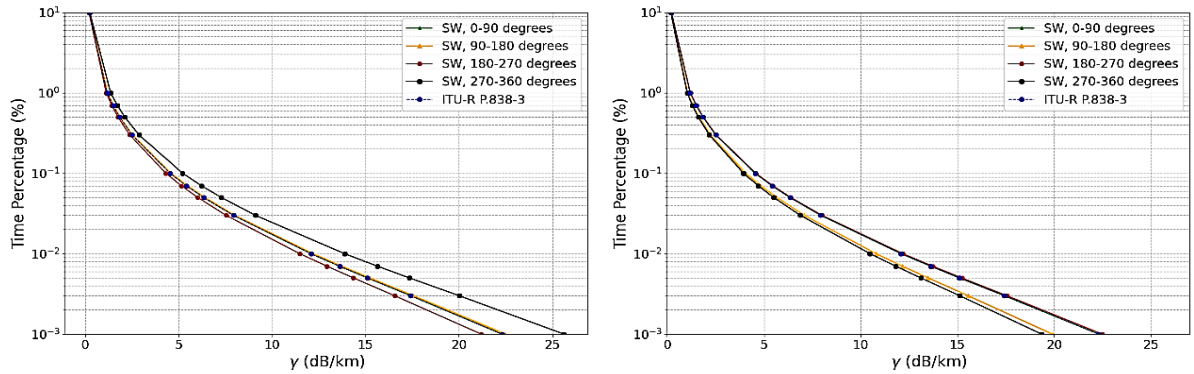
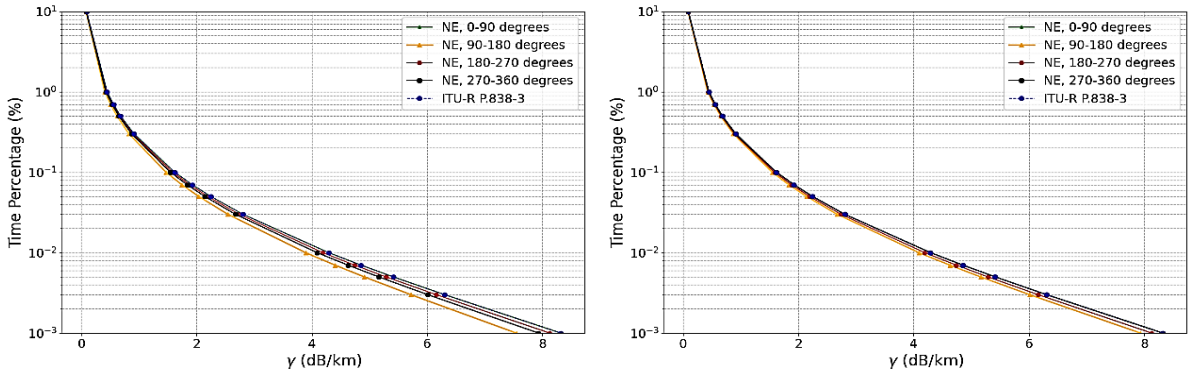


Figure 2. CCDF of specific attenuation (γ, dB/km) during the SW monsoon for 40–50 mm/h against ITU-R P.838-3 at 38 GHz under (a) unconstrained winds and (b) constrained winds (9–11 m/s)

During the SW monsoon, the CCDFs in Figure. 1 and 2 exhibit the widest directional spread and the clearest separation from the ITU-R curve, particularly at 38 GHz. This indicates that SW conditions produce stronger sensitivity to DSD variability and wind-driven microphysical effects, consistent with convective dominance in tropical rainfall [20]. Similar behavior has been reported for 32/38 GHz terrestrial links where rare-event attenuation exceeds theoretical predictions, suggesting the importance of local microphysics in model mismatch [12]. Local assessments in equatorial Malaysia also highlight persistent deviations even with recent ITU-R updates, reinforcing the need for local characterization in such climates [10]. The quantitative error statistics confirm this CCDF observation: at 38 GHz under unconstrained winds, the 270°–360° sector exhibits the largest mismatch, with RMSE = 1.550 dB/km and

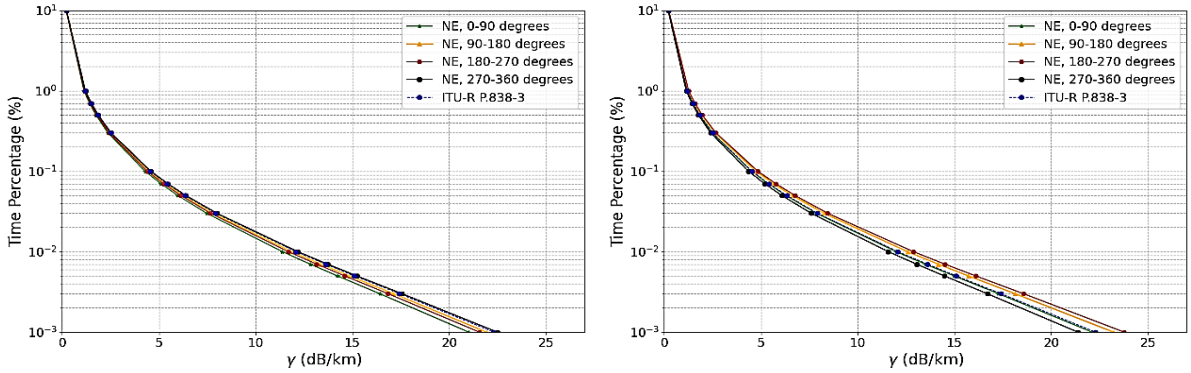
bias = +1.193 dB/km (Table 4), indicating an ITU-R underestimation. Importantly, the bias is not uniform across directions; the 180°–270° sector shows the opposite tendency (bias = −0.393 dB/km, Table 4), meaning ITU-R can also overestimate depending on the wind sector. Under constrained winds (9–11 m/s), the deviation remains high in the 270°–360° sector (RMSE = 1.383 dB/km, Table 5). Overall, SW is the season with the most pronounced directional dependence, and reliance on a single rainfall-rate-based model can lead to the greatest planning uncertainty at 38 GHz.

In contrast, the NE monsoon CCDFs in Figures 3 and 4 remain close to ITU-R P.838-3 across most sectors under both wind classifications. This implies reduced direction-dependent variability compared with SW and suggests more stable attenuation behavior during NE. The error statistics support this: the 38 GHz RMSE remains comparatively modest, with the worst-case value staying below about 0.6 dB/km under unconstrained winds (Table 4), and bias values generally remain small in magnitude (Tables 4 and 5). In practice, this means ITU-R P.838-3 is more likely to provide an acceptable baseline for NE-season planning, with fewer sectors exhibiting large systematic offsets.



(a) (b)

Figure 3. CCDF of specific attenuation (γ, dB/km) during the NE monsoon for 40–50 mm/h against ITU-R P.838-3 at 18 GHz under (a) unconstrained winds and (b) constrained winds (9–11 m/s)

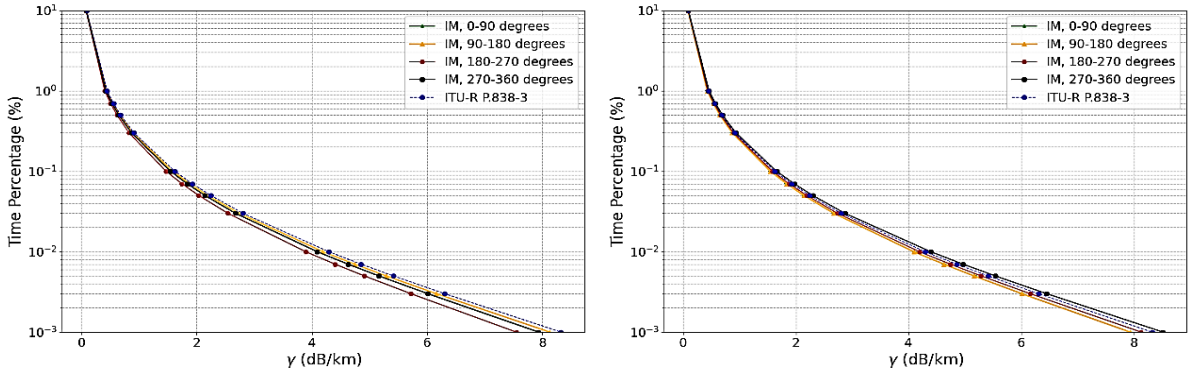


(a) (b)

Figure 4. CCDF of specific attenuation (γ, dB/km) during the NE monsoon for 40–50 mm/h against ITU-R P.838-3 at 38 GHz under (a) unconstrained winds and (b) constrained winds (9–11 m/s)

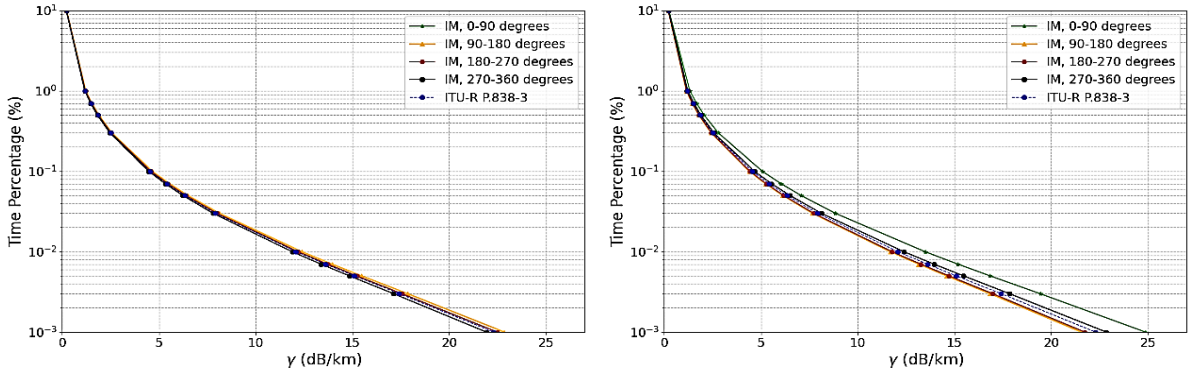
The IM season exhibits transitional behavior. The CCDFs in Figures 5 and 6 remain close to ITU-R at 18 GHz, but at 38 GHz, a clearer separation between sectors emerges, especially under constrained winds. This indicates that wind-direction sensitivity can reappear during the

IM, consistent with mixed meteorological conditions that can shift DSD characteristics between events. The constrained-wind statistics in Table 5 capture this directional dependence: the 0°–90° sector has an RMSE of 1.212 dB/km and a bias of +0.936 dB/km, indicating underestimation by ITU-R, whereas the 90°–180° sector shows overestimation (bias = −0.257 dB/km). The presence of both positive and negative bias across directions confirms that the ITU-R error direction is not fixed during the IM. This aligns with prior findings that local rainfall regimes strongly influence DSD behavior and specific-attenuation exceedance statistics [23], and that model errors become more pronounced at mm-wave frequencies when evaluating very-low-time percentages [3, 18].



(a) (b)

Figure 5. CCDF of specific attenuation (γ, dB/km) during the IM monsoon for 40–50 mm/h against ITU-R P.838-3 at 18 GHz under (a) unconstrained winds and (b) constrained winds (9–11 m/s)



(a) (b)

Figure 6. CCDF of specific attenuation (γ, dB/km) during the IM monsoon for 40–50 mm/h against ITU-R P.838-3 at 38 GHz under (a) unconstrained winds and (b) constrained winds (9–11 m/s)

From a system-design perspective, the practical significance of these findings is tied to the exceedance levels used in link planning. The 0.01% exceedance level (approximately 52.6 min/year under one-minute sampling) is commonly used as a reference for fade-margin planning, while 0.001% (about 5.26 min/year) becomes relevant for more stringent availability targets. The CCDF results show that differences are generally small at 18 GHz, supporting ITU-R as a practical reference at this band. At 38 GHz, however, the largest departures occur under SW and IM in specific wind-direction sectors, precisely where accurate estimation is most critical for reliable operation.

Table 4. RMSE and bias between the DSD-derived specific attenuation and ITU-R P.838-3 at 18 and 38 GHz for 40–50 mm/h under unconstrained winds.

Monsoon	Wind directions [°]	Specific attenuation at 18 GHz [dB/km]		Specific attenuation at 38 GHz [dB/km]	
		RMSE	Bias	RMSE	Bias
SW	**0°–90°**	0.183	-0.136	0.000	0.000
	90°-180°	0.000	0.000	0.107	+0.086
	180°-270°	0.269	+0.207	0.518	−0.393
	270°-360°	0.191	+0.150	1.550	+1.193
NE	**0°–90°**	0.000	0.000	0.601	−0.450
	90°-180°	0.365	-0.279	0.179	−0.121
	180°-270°	0.096	-0.064	0.336	−0.257
	270°-360°	0.183	-0.136	0.107	+0.086
IM	**0°–90°**	0.183	-0.136	0.107	+0.086
	90°-180°	0.096	-0.064	0.254	+0.200
	180°-270°	0.365	-0.279	0.107	+0.086
	270°-360°	0.183	-0.136	0.179	−0.121

A practical example from Malaysia illustrates this relevance, where a 38 GHz system is designed to operate with a specific attenuation of 18.4 dB/km during monsoon conditions [7]. Consistent with the broader literature, prediction errors tend to worsen around 0.01% and 0.001% exceedance levels [1], [17], and underprediction at 30–300 GHz has been reported, particularly below 0.01% [3]. Therefore, the key implication is that fade-margin allocation and link-orientation decisions at 38 GHz in tropical climates should not rely solely on ITU-R; seasonal phase and wind-direction variability can meaningfully affect both the magnitude and the sign of the prediction error.

Table 5. RMSE and bias between the DSD-derived specific attenuation and ITU-R P.838-3 at 18 and 38 GHz for 40–50 mm/h under constrained winds (9–11 m/s)

Monsoon	Wind directions [°]	Specific attenuation at 18 GHz [dB/km]		Specific attenuation at 38 GHz [dB/km]	
		RMSE	Bias	RMSE	Bias
SW	**0°–90°**	0.183	-0.136	0.000	0.000
	90°-180°	0.541	+0.407	1.120	−0.850
	180°-270°	0.269	+0.207	0.107	+0.086
	270°-360°	0.183	-0.136	1.383	−1.050
NE	**0°–90°**	0.000	0.000	0.076	−0.043
	90°-180°	0.183	-0.136	0.433	+0.336
	180°-270°	0.096	-0.064	0.707	+0.550
	270°-360°	0.000	0.000	0.421	−0.314
IM	**0°–90°**	0.183	-0.136	1.212	+0.936
	90°-180°	0.183	-0.136	0.336	−0.257
	180°-270°	0.096	-0.064	0.259	−0.186
	270°-360°	0.107	+0.086	0.254	+0.200

The results indicate that ITU-R P.838-3 performance in a tropical setting is condition-dependent: it is generally adequate at 18 GHz and more reliable during the NE monsoon, but

shows larger, direction-dependent deviations at 38 GHz, especially during the SW and IM periods. For practical 5G/6G mm-wave deployment, this supports using local DSD-derived statistics to (i) identify worst-case wind-direction sectors, (ii) adjust fade margins for the most sensitive seasonal conditions, and (iii) guide link orientation and robustness measures, including alignment tolerance and mounting stability, to reduce outage risk under severe low-exceedance conditions. Importantly, the deviation metrics in Tables 4 and 5 show that the mismatch at 38 GHz can reach RMSE values on the order of 1–1.6 dB/km in the most affected sectors, with the bias sign changing across wind directions, meaning ITU-R may either under- or overestimate depending on the sector. Such sector-dependent behavior cannot be captured by frequency alone and suggests that a rainfall-rate-only framework is insufficient when seasonal variability, wind direction, and wind-driven DSD changes are present. For current 5G links, these deviations may be manageable through conservative margin allocation and practical mitigation, such as tighter alignment tolerance and adaptive fade management. However, as availability targets become more stringent, low-exceedance behavior becomes increasingly critical [27], and adopting monsoon-aware, direction-dependent, and adaptive parameterization will be pivotal to sustaining link performance during fluctuating weather conditions.

4. CONCLUSIONS

This study indicates that the current ITU-R P.838-3 standard does not fully account for the propagation characteristics of tropical rainfall at mm-wave frequencies, which are critical for 5G and future 6G networks. Findings presented in this paper indicate that the greatest discrepancy occurs at 38 GHz during the SW monsoon season, as unconstrained winds lead to excessive specific attenuation relative to the ITU-R predicted values. The 270°–360° sector records the highest deviation under unconstrained winds, with RMSE = 1.550 dB/km and bias = +1.193 dB/km, indicating that ITU-R underestimates the specific attenuation. Under constrained winds (9–11 m/s), the deviation remains high (RMSE = 1.383 dB/km), and the bias becomes sector-dependent. Conversely, during constrained winds, the predicted values are overestimated. During the IM season, there are moderate discrepancies, but they remain sensitive to direction; whereas during the NE monsoon season, the ITU-R values are largely consistent. At 18 GHz, across all seasons and wind conditions, the deviations remain small, with RMSE generally below about 0.6 dB/km and bias values close to zero in all seasons and wind conditions, demonstrating that at this mm-wave band, the ITU-R models could be applied without hassle.

These findings show the effects of factors other than rain rates on specific attenuation, such as seasonal variability, wind direction, and variations in DSD, which introduce complex spatial and temporal dependencies beyond what frequency-only models can account for. The presented results could provide the necessary adaptive requirements for the development of a monsoon-aware, direction-specific fade-margin framework for rainfall prediction models, to warrant high capacity and fulfill stringent availability requirements for 5G and 6G mm-wave links.

ACKNOWLEDGEMENT

This work was conducted as part of the author's doctoral research at the Faculty of Electrical Engineering, Universiti Teknologi Malaysia (UTM), with collaboration from Universiti Tun Hussein Onn Malaysia (UTHM). The author also acknowledges the Ministry of Higher Education (MOHE) and the Department of Polytechnic and Community College Education (DPCCE) for their continuous support. This work received no external research funding.

REFERENCES

[1] S. Shrestha and D.-Y. Choi, "Rain attenuation over terrestrial microwave links in South Korea," Iet Microwaves Antennas & Propagation, vol. 11, no. 7, pp. 1031–1039, Jun. 2017, doi: 10.1049/IET-MAP.2016.0553.

[2] J. Tan and M. Thurai, "Rain-induced cross polarisation on line-of-sight systems at 38 GHz," vol. 144, no. 1, pp. 20–26, Feb. 1997, doi: 10.1049/IP-MAP:19970962.

[3] I. Mata-Alonso, J. M. Riera, L. Luini, H. Y. Lam, and D. Pimienta-del-Valle, "Rain Attenuation at Millimeter Waves in Different Climatic Zones Estimated from Drop Size Distributions," pp. 1–5, Mar. 2024, doi: 10.23919/eucap60739.2024.10501081.

[4] M. Rashid, J. Din and O. A. Aziz, " Spatial variations of rain intensity over a short length propagation for 5G links based on a rain gauge network", doi: 0.12928/TELKOMNIKA.v19i2.16809.

[5] H. Y. Lam, J. Din, and S. L. Jong, "Statistical and Physical Descriptions of Raindrop Size Distributions in Equatorial Malaysia from Disdrometer Observations," Advances in Meteorology, vol. 2015, pp. 1–14, 2015, doi: https://doi.org/10.1155/2015/25373.

[6] H. Y. Lam, L. Luini, J. Din, M. J. Alhilali, S. L. Long, and F. Cuervo, "Impact of rain attenuation on 5G millimeter-wave systems in equatorial Malaysia investigated through disdrometer data," in Proc. EuCAP 2017, pp. 1793-1707, 2017, doi: 10.23919/EuCAP.2017.7928616.

[7] I. A. Shayea, S. A. Alkhawaldeh, M. A. Nisirat, T. A. Rahman, and M. Ergen, "Channel fading attenuation based on rainfall rate for future 5G wireless communication system over 38-GHz," International Journal of Electrical and Computer Engineering (IJECE), vol. 12, no. 5, p. 5104, Oct. 2022, doi: https://doi.org/10.11591/ijece.v12i5.pp5104-5113.

[8] H. Y. Lam, L. Luini, J. Din, C. Capsoni, and A. D. Panagopoulos, "Investigation of Rain Attenuation in Equatorial Kuala Lumpur," IEEE Antennas and Wireless Propagation Letters, vol. 11, pp. 1002–1005, 2012, doi: https://doi.org/10.1109/lawp.2012.221437.

[9] Asma Ali Budalal et al., "Effective Rain Rate Model for Analysing Overestimated Rain Fade in Short Millimetre-Wave Terrestrial Links Due to Distance Factor," Results in Engineering, pp. 104112–104112, Jan. 2025, doi: https://doi.org/10.1016/j.rineng.2025.104112.

[10] F. A. Semire, R. Mohd-Mokhtar, and I. A. Akanbi, "Validation of New ITU-R Rain Attenuation Prediction Model over Malaysia Equatorial Region," MAPAN, vol. 34, no. 1, pp. 71–77, Nov. 2018, doi: https://doi.org/10.1007/s12647-018-0295-z.

[11] H Ju, Y Zou and G.M. Rebeiz, " A 6–19 GHz Reconfigurable I/Q Receiver with 21 dB Gain, 3 dB NF, and 30 dB IRR for 6G FR3 in 22-nm FD-SOI," 2025 IEEE BiCMOS and Compound Semiconductor Integrated Circuits and Technology Symposium (BCICTS), San Diego, USA, 2025, pp. 1-4, doi: https://doi.org/10.1155/2019/1712791.

[12] B. B. Jimoh, A. Y. Abdulrahman, A. J. Falade, O. Oniyide, S. O. Zakariyya, and T. A. Rahman, "Evaluation of the frequency scaling prediction techniques using experimental data," Telecommunications and Radio Engineering, vol. 76, no. 5, pp. 433–442, Jan. 2017, doi: https://doi.org/10.1615/telecomradeng.v76.i5.40.

[13] M. Woldamanuel and F. D. Diba, "Enhanced adaptive code modulation for rainfall fade mitigation in Ethiopia," EURASIP Journal on Wireless Communications and Networking, vol. 2022, no. 1, Jan. 2022, doi: https://doi.org/10.1186/s13638-021-02085-0.

[14] P. Valtr, M. Fencl, V. Bareš and P. Pechač, "Comparison of measured and theoretically predicted rain attenuation at 32 GHz and 38 GHz," 12th European Conference on Antennas and Propagation (EuCAP 2018), London, UK, 2018, pp. 1-3, doi: 10.1049/cp.2018.0377.

[15] M. R. Islam, J. Chebil, O. O. Khalifa, S. Khan, H. Dao and A. -H. Zyoud, "Effect of frequency on fade slope based on rain attenuation data measured in malaysia," International Conference on Computer and Communication Engineering (ICCCE'10), Kuala Lumpur, Malaysia, 2010, pp. 1-4, doi: 10.1109/ICCCE.2010.5556834.

[16] F. Capelletti and L. Luini, "An Effective Approach to Instantaneous Rain Attenuation Frequency Scaling Using Single or Multiple Satellite-Based Measurements," in IEEE Transactions on Antennas and Propagation, vol. 73, no. 7, pp. 4853-4862, July 2025, doi: 10.1109/TAP.2025.3552227.

[17] M. Ghanim, M. Alhilali, J. Din, and H. Y. Lam, "Rain attenuation statistics over 5G millimetre wave links in Malaysia," Indonesian Journal of Electrical Engineering and Computer Science, vol. 14, no. 2, p. 1012, May 2019, doi: https://doi.org/10.11591/ijeecs.v14.i2.pp1012-1017.

[18] A. Kumar Verma, R. Nandan and A. Verma, "Regional variability of Specific rain attenuation in mm wave region for Indian Sub-tropical climate," 2018 IEEE Indian Conference on Antennas and Propogation (InCAP), Hyderabad, India, 2018, pp. 1-4, doi: 10.1109/INCAP.2018.8770886.

[19] A. M. Al-Saegh, A. Sali, J.S. Mandeep, and A. Ismail, "Extracted atmospheric impairments on earth-sky signal quality in tropical regions at Ku-band," Journal of Atmospheric and Solar-Terrestrial Physics, vol. 104, pp. 96–105, Sep. 2013, doi: https://doi.org/10.1016/j.jastp.2013.08.018.

[20] A. Aresu, P. Martellucci and P. Migliorini, "Depolarizing properties of rain, measured on a 30 GHz short path terrestrial link," International Symposium on Antennas and Propagation Society, Merging Technologies for the 90's, Dallas, TX, USA, 1990, pp. 1824-1827 vol.4, doi: 10.1109/APS.1990.115486.

[21] A. M. Al-Saegh, A. Sali, M. Singh, A. Ismail, and A. H. J. Aljumaily, "Earth-sky link quality performance for fixed and mobile scenarios in tropical regions," Progress In Electromagnetics Research C, vol. 39, pp. 61–75, 2013, doi: https://doi.org/10.2528/pierc13022703.

[22] N. N. Md Yusof, J. Din, and H. Y. Lam, "Enhancing millimeter-wave communication: a tropical perspective on raindrop size distribution and signal attenuation," International Journal of Power Electronics and Drive Systems, vol. 15, no. 1, p. 467, Nov. 2024, doi: 10.11591/ijece.v15i1.pp467-479

[23] Y. Li, Y. Zheng, G. Chen, L. Li and F. Xu, "Difference analysis of raindrop size distribution characteristics in Meiyu and typhoon types of short-term heavy rainfall in summer in Jiangsu Province," Transactions of Atmospheric Sciences, vol. 47, no. 5, pp. 798-808, 2024, doi: 10.13878/j.cnki.dqkxxb.20240422002

[24] M. Rashid, J. Din, H. Y. Lam and O. A. Aziz, "Preliminary Investigation of Small Scale Spatial Variability of Rain Intensity Using a Rain Gauge Network for Mm-Wave Propagation Application," *2021 IEEE Asia-Pacific Conference on Applied Electromagnetics (APACE)*, Penang, Malaysia, 2021, pp. 1-5, doi: 10.1109/APACE53143.2021.9760672

[25] Z. Saberi, A. Fudholi and K. Sopian, "Fitting of weibull distribution method to analysis wind energy potential at Kuala Terengganu, Malaysia," Journal of Advanced Research in Fluid Mechanics and Thermal Sciences, vol. 66, no. 1, pp. 1-11, Feb 2020,

[26] Y. A. Ahmad, A. F. Ismail, A. Hamid, and M. F. Jamlos, "Area-Based Rainfall Rate Model for Specific Attenuation in the Equatorial Region," IIUM Engineering Journal, vol. 25, no. 2, pp. 287–298, Jul. 2024, doi: https://doi.org/10.31436/iiumej.v25i2.3279

[27] A. Shukur, Y. A. Ahmad, Muhammad, and Khairayu Badron, "Latency Performance Evaluation of LEO Starlink and SES-12 GEO HTS Network Under Tropical Rainfall Conditions," IIUM Engineering Journal, vol. 26, no. 2, pp. 204–219, May 2025, doi: https://doi.org/10.31436/iiumej.v26i2.3653.

https://doi.org/10.31436/iiumej.v27i2.4059

MultiModal Explainable AI and Blockchain Integration for Automated Halal Verification of Cosmetic Products

ENDANG SUPRIYATI[1*], MOHAMMAD IQBAL[2], TRI LISTYORINI[1]

[1]*Dept. of Informatics Engineering, Universitas Muria Kudus, Kudus, Indonesia*
[2]*Dept. of Electrical Engineering, Universitas Muria Kudus, Kudus, Indonesia*

**Corresponding author: endang.supriyati@umk.ac.id*

(Received: 23 November 2025; Accepted: 12 February 2026; Published online: 10 May 2026)

ABSTRACT: This research develops a framework that integrates blockchain technology and MXAI (Multimodal Explainable Artificial Intelligence) to automate the authentication and verification of halal cosmetic products. Two data sources were used: text extracted by OCR (Optical Character Recognition) and text manually input from cosmetic labels. After the pre-processing stage, the text data are analyzed using zero-shot classification to determine the inspection results, namely halal, haram, or syubhat. The inspection is conducted using MXAI, with decisions based on confidence scores and SHAP values. The inspection results are converted into digital reports as SHA-256 hashes and stored as Merkle roots on the blockchain, allowing users to download certificates as QR codes. The halal status experiment on the cosmetics dataset achieved an accuracy of 94.5% for classification, with a 3.3% improvement over baseline models, evaluated using stratified 5-fold cross-validation. This system enhances transparency, accountability, and public trust in automated halal certification. The contribution of this research is the integration of MXAI and blockchain technology into a single intelligent halal verification system, which can be extended to other supply chain sectors.

ABSTRAK: Kajian ini menggabungkan teknologi rantaian blok dan MXAI (Kecerdasan Buatan Penjelasan Multimodal) bagi automasi dan jaminan pengesahan produk kosmetik halal. Dua sumber data digunakan: teks yang diekstrak oleh OCR (Pengesahan Optik Karakter) dan teks yang dimasukkan secara manual daripada label kosmetik. Selepas peringkat pra-pemprosesan, data teks dianalisis menggunakan klasifikasi sifar-tembakan bagi menentukan keputusan pemeriksaan, iaitu halal, haram, atau syubhat. Pemeriksaan dijalankan menggunakan MXAI, dengan keputusan berdasarkan skor keyakinan dan nilai SHAP. Keputusan pemeriksaan ditukar menjadi laporan digital dalam bentuk hash SHA-256 dan disimpan sebagai punca Merkle pada rantaian blok, membolehkan pengguna memuat turun sijil dalam bentuk kod QR. Eksperimen status halal pada set data kosmetik menunjukkan ketepatan 94.5% untuk prestasi pengelasan, dengan peningkatan 3.3% berbanding model asas, dinilai menggunakan pengesahan silang 5-lipatan berstrata. Sistem ini meningkatkan ketelusan, akauntabiliti, dan kepercayaan awam dalam pensijilan halal automatik. Sumbangan penyelidikan ini ialah melalui integrasi MXAI dan teknologi rantaian blok dalam satu sistem pengesahan halal pintar tunggal, yang boleh diperluas kepada sektor rantaian bekalan lain.

KEYWORDS: *Halal Verification System, Blockchain, MXAI (Multimodal Explainable Artificial Intelligence), Multimodal Learning, OCR (Optical Character Recognition).*

https://doi.org/10.31436/iiumej.v27i2.4059

1. INTRODUCTION

Artificial Intelligence (AI) has become increasingly integrated into modern society and has become an important component of it since the Industrial Revolution 4.0. AI has been used in various sectors. The accuracy of AI-driven decision-making and automation makes it an irreplaceable tool. The use of AI has had an impact on various industrial sectors, including the halal cosmetics industry, due to consumer demand for high-quality halal products[1].

Research on the halal cosmetics industry remains limited, particularly regarding the use of artificial intelligence (AI). Research in this field still faces several challenges, such as raw material analysis, halal certification processes, and marketing strategies for both Muslim and non-Muslim markets. The development of AI has had a significant impact on various sectors, including the optimization of halal product development and certification procedures [2].

Artificial Intelligence (AI) capabilities are far superior to traditional decision support systems in processing large, complex datasets, especially in making more accurate and efficient decisions. The integration of real-time decision-making with intelligent automation is replacing traditional decision support systems [3]. The ability to learn from new data and adapt to emerging challenges is one of AI's remarkable capabilities. This makes it particularly important for industries that require quick, accurate, and timely decisions.

At the same time, blockchain technology is reshaping various sectors such as healthcare, supply chain management, and other fields by enhancing data security and maintaining integrity through permanent audit trails [4]. In the halal supply chain, blockchain ensures traceability and transparency, preventing manipulation and maintaining the authenticity of halal products [5], [6]. The decentralization of blockchain makes it crucial for securing the halal certification process and fostering trust within the industry.

The integration of blockchain and AI has proven transformative across various industrial sectors. In many higher education institutions, this integration enables efficient certificate management by automating verification through smart contracts, thereby enhancing security and transparency [7], [8]. This synergy offers innovative applications in identity verification, authenticity detection, and contract management [9], [10]. Similarly, the industry benefits from an efficient and more reliable certification system that utilizes both technologies [11], [12].

The halal certification process remains inefficient despite being essential for Muslims. The certification process is time-consuming, costly, and prone to fraud and mislabeling. Regulatory challenges and inconsistencies complicate global halal standards [13], [14] Thus, there is an urgent need to digitalize the certification process to ensure data authenticity and efficiency.

Although AI and blockchain have been applied in the food industry and supply chain management [15],[4]. There is still no integrated framework that combines the two, so there is a need to integrate blockchain with Explainable Artificial Intelligence (XAI), especially in the cosmetics industry. In the cosmetics industry, challenges such as ingredient analysis, system transparency, and certification compliance remain unaddressed. This research aims to develop an integrated framework that combines XAI and blockchain to automate and improve the halal verification process for cosmetics.

This research proposes a new approach that combines Multimodal Explainable AI (MXAI) with blockchain technology to verify the halal status of cosmetic products. The system will analyze product label images, ingredient texts, and supporting documents. It will

offer explainable interpretations of halal classification. With the potential of AI and blockchain, this framework will provide a more efficient, transparent, and secure halal certification process. It will contribute to the modernization of halal governance and meet the demand for trustworthy halal products.

2. METHODOLOGY

2.1. Research Design

This research uses a prototype system development design based on MXAI (Multimodal Explainable Artificial Intelligence) and Blockchain to verify the halal status of cosmetic products. The data used comes from two sources: packaging label images and ingredient composition texts. The MXAI method was chosen because it allows the system to understand visual and textual contexts simultaneously, while blockchain technology ensures the security, transparency, and traceability of verification results. The research stages included: 1. collection of cosmetic label datasets, 2. multimodal data pre-processing, 3. MXAI model training and testing, 4. blockchain integration for digital recording and auditing, and 5. system testing through a Gradio-based user interface (UI). The overall architecture of the proposal, MXAI and blockchain-based halal verification framework, is illustrated in Fig. 1.

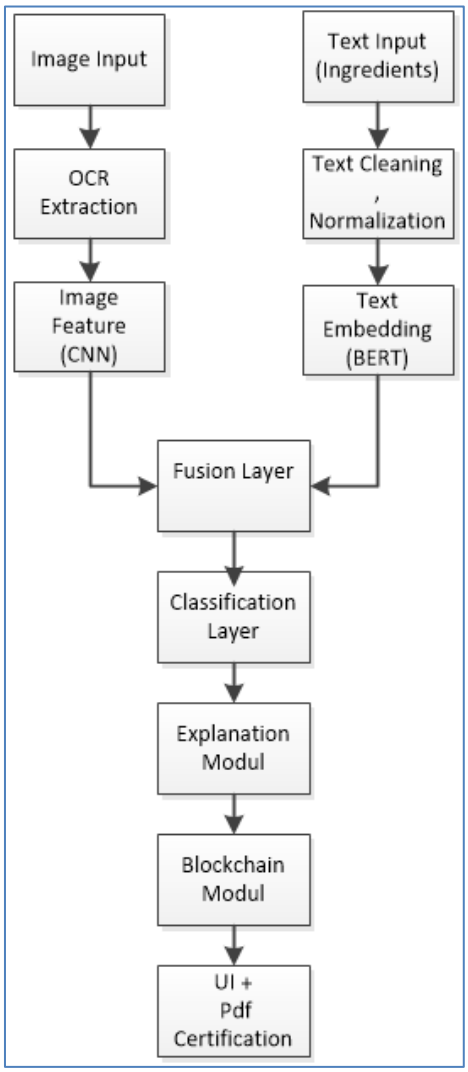


Figure 1. Halal verification framework for cosmetic products

https://doi.org/10.31436/iiumej.v27i2.4059

2.2. Data Source

The data used consisted of two main sources, textual data and visual data:

1) *Textual Data*: data in the form of a list of cosmetic ingredients that is manually entered by users into the system. The data format is text that lists product ingredients, such as water, tocopherol, and glycerine. This data is used as a reference for the validation and classification process of the halal status of ingredients.
2) *Data Visualization*: data in the form of cosmetic packaging labels containing ingredient composition information. The images are processed with Optical Character Recognition (OCR), and the text is extracted using the *pytesseract* library. The extracted text is then compared and fused with manually inputted textual data to produce a combined representation.

Both types of data are processed in parallel using multimodal learning. Information from text and images is combined using attention-based embedding fusion techniques to improve the accuracy of classifying a cosmetic product as halal, dubious, or haram.

2.3. Data Preprocessing and Workflow Procedure

The data processing stages in this study were carried out in stages. Data obtained from text and OCR was processed by the MXAI model. The data processing steps included:

2.3.1. Manual data input and OCR Labeling

The text of the cosmetic ingredient composition is entered into the system, or the label image is uploaded. The system will extract the text from the uploaded image data using Optical Character Recognition (OCR) with pytesseract.

2.3.2. Text Cleaning and Normalization

The text results from OCR and manual input are then cleaned of noise (punctuation marks, capital letters, and ASCII characters) using regular expressions. All text is converted to lowercase to maintain consistency.

2.3.3. Stopword Removal and Tokenization

Words that do not have significant meaning (not, in, to, from, and others) are removed using Indonesian and English stopword lists. Tokenization is the process of converting words into their root forms.

2.3.4. Word Embedding

The cleaned text is converted into a numerical representation using the pre-trained multilingual BERT-base language model, and then contextual embeddings are generated for each word in the material.

2.3.5. Feature Extraction

For image data, product label images are resized to 224×224 pixels, normalized, and then fed into a Convolutional Neural Network (CNN).

2.3.6. Feature Fusion Layer

Representations of two data sets (text and image) are combined using a weighted fusion layer.

$$\mathbf{s} = \alpha \mathbf{s}_{text} + (1 - \alpha)\mathbf{s}_{image}, \quad 0 \leq \alpha \leq 1 \tag{1}$$

https://doi.org/10.31436/iiumej.v27i2.4059

where $\mathbf{s}_{text}$ is the class-score vector produced from the text encoder, $\mathbf{s}_{image}$ is the class-score vector produced from the image encoder, and α is the modality-weighting coefficient learned or tuned during training.

2.3.7. Classification and Explainability

The resulting combined score is sent to the classification layer, which produces the final label (halal, dubious, haram). The system activates the explainability module to highlight the words that most influence the model's decision.

2.3.8. Blockchain recording

The verification results, which include the halal status, document hash, and timestamp, are recorded in the blockchain ledger using the SHA-256 algorithm. Each block contains the following structure:

$$B_i = \{i, t_i, d_i, h_{i-1}, h_i\} \tag{2}$$

where i is the block index, t_i is the timestamp, d_i is the verification data, h_{i-1} is the previous block hash, and h_i is the current block hash.

2.4. MXAI Architecture

The MXAI (Multimodal Explainable AI) architecture has three main components: the text encoder, the image encoder, and the Fusion and Classification Layer. These three components work in an integrated manner to produce predictions of halal, dubious, or haram status based on multimodal analysis.

2.4.1. Text Encoder (BERT)

The text encoder uses Bidirectional Encoder Representations from Transformers (BERT) to capture semantic relationships among ingredient terms. Each cosmetic ingredient sequence is represented as:

$$T = \{t_1, t_2, t_3, \dots, t_n\}$$

The contextual text representation is then obtained as:

$$\mathbf{h}_{text} = f_{BERT}(T) \tag{3}$$

where T denotes the tokenized ingredient sequence, t_i represents the i-th ingredient token, and $\mathbf{h}_{text}$ is the contextual vector representation of the cosmetic product ingredient text, e.g., the relationship between alcohol, glycerine, and collagen, for further analysis by the AI model.

2.4.2. Image Encoder (CNN)

This component extracts visual features from cosmetic product packaging labels. OCR processes images I are processed through a Convolutional Neural Network to extract visual semantic features

$$\mathbf{h}_{image} = f_{CNN}(I) \tag{4}$$

where I is the input product-label image and $\mathbf{h}_{image}$ is the visual feature representation extracted by the CNN.

2.4.3. Layer Fusion and Classification Features

After the representations of the text and image are obtained, they are combined using a weighted fusion layer with the formula:

https://doi.org/10.31436/iiumej.v27i2.4059

$$\mathbf{s} = \alpha \mathbf{s}_{text} + (1 - \alpha)\mathbf{s}_{image},\ 0 \leq \alpha \leq 1 \quad (5)$$

Where $\mathbf{s}$ is the fused class-probability vector, $\mathbf{s}_{text}$ is the probability score of text analysis results, $\mathbf{s}_{image}$ is the image analysis probability score, and α is the weight of the text modality contribution, adjusted during the training process to achieve optimal results

The final value $\mathbf{s}$ is used to determine the halal classification of products with decision functions:

$$\hat{y} = \arg\max_{c \in \mathcal{C}} \mathbf{s}_c \quad (6)$$

where $\mathcal{C} = \{halal, syubhat, haram\}$, $\mathbf{s}_c$ is the predicted score for the class c, and $\hat{y}$ is the final predicted halal-status label.

2.4.4. Loss Function

In addition, the model was trained using a combined loss function that optimized accuracy while maintaining consistency across modalities. The total loss function was defined as follows:

$$\mathcal{L}_{total} = \mathcal{L}_{CE} + \lambda \mathcal{L}_{InfoNCE} \quad (7)$$

$$\mathcal{L}_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{\mathcal{C}} y_{i,c} \log\left(\hat{p}_{i,c}\right) \quad (8)$$

$$\mathcal{L}_{InfoNCE} = -\log \frac{\exp\left(\frac{sim\left(\mathbf{h}_{text}, \mathbf{h}_{image}\right)}{\tau}\right)}{\sum_{j=1}^{N} \exp\left(\frac{sim\left(\mathbf{h}_{text}, \mathbf{h}_{image,j}\right)}{\tau}\right)} \quad (9)$$

where $\mathcal{L}_{CE}$ is the Cross-Entropy Loss for multi-class classification, $\mathcal{L}_{InfoNCE}$ is the contrastive alignment loss that ensures consistency between text and images, λ is the coefficient regulating the contribution between the two losses, and τ is the temperature parameter that controls the sharpness of similarity scores in the contrastive loss.

The methods ensure that the system is not only capable of recognizing halal status from textual and visual content but also has strong cross-modal alignment, resulting in more accurate and explainable decisions.

2.4.5. Blockchain Recording

For traceability, each verification record is stored in a Python-based blockchain. Each block structure is defined as described in Eq. (2). Then, the hash block i calculated using the SHA-256 algorithm:

$$h_i = SHA256(i \parallel t_i \parallel d_i \parallel h_{i-1}) \quad (10)$$

where ‖ denotes concatenation.

2.4.6. Merkle Root Verification

To ensure data integrity, all transactions in a block are summarized using a Merkle tree, with the formula:

$$R = H\big(H(T_1 \parallel T_2) \parallel H(T_3 \parallel T_4)\big) \quad (11)$$

where R is the Merkle root, $H(\cdot)$ denotes the SHA-256 hash function, T_1, T_2, T_3, T_4 are transaction records, and ‖ denotes concatenation.

2.5. Implementation Tool

The test environment and implementation tools used in this study are summarized in Table 1. All components were selected to ensure compatibility between AI, blockchain, and document module creation. This configuration is designed to support multimodal data processing, model deployment, and certificate creation within the verification integration framework.

Table 1. The test environment and implementation tools used in this study

Component	Technology/Tool
Programming Language	Python 3.10
AI Framework	PyTorch, Transformers
Supporting Libraries	Pytesseract, OpenCV, Gradio, qrcode, ReportLab, haslib
Blockchain Framework	Custom lightweight blockchain
Development Environment	Google Colab, Gradio
Certificate Format	PDF certificate with QR code generated using ReportLab and qrcode

The configuration of tools in Table 1 optimizes system performance during model training, blockchain verification, and certificate creation. The integration of Gradio and ReportLab enables interactive, transparent prototypes for halal product verification, while the use of Python-based blockchain-lightweight supports efficient ledger recording and data integrity validation.

2.6. Data Analysis Techniques

The data analysis stage was conducted to evaluate the performance of the MXAI and Blockchain verification system. Two types of analysis were used: (1) AI model performance and (2) Blockchain integrity.

2.6.1. AI Model Measurement

Classification accuracy using the standard formula is:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{12}$$

The measurements are True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), respectively. Precision and recall measurements are also provided, and the F1 score is calculated to assess class-balanced performance.

$$Precision = \frac{TP}{TP+FP}, Recall = \frac{TP}{TP+FN}, F_1 = \frac{2\times Precision\times Recall}{Precision+Recall} \tag{13}$$

The consistency of explanations is measured using the metrics of fidelity and sparsity, defined as follows:

$$Fidelity = 1 - \frac{|f(x)-f(x_{\setminus S})|}{|f(x)|+\epsilon}, Sparsity = \frac{\|S\|_0}{\|x\|_0} \tag{14}$$

where $f(x)$ is the model output for the original input x, $x_{\setminus S}$ is the input after removing the important features identified by SHAP, S is the set of features selected by SHAP, $\|\cdot\|_0$ denotes the number of non-zero elements, and ϵ is a small constant used to avoid division by zero. High fidelity and low sparsity indicate that the explanation is faithful to the model decision while remaining concise.

https://doi.org/10.31436/iiumej.v27i2.4059

2.6.2. Blockchain integrity.

Blockchain integrity was validated using hash consistency across linked data blocks, defined as:

$$H_i = SHA256(i \parallel t_i \parallel d_i \parallel H_{i-1}) \quad (15)$$

where i is the block index, t_i is the timestamp, d_i is the verification data, H_{i-1} is the hash of the previous block, and $\parallel$ denotes concatenation.

The verification is successful when:

$$H_i^{computed} = H_i^{recorded} \quad (16)$$

This ensures that no unauthorized modification occurs in the current or previous linked blocks. Data analysis, therefore, supports accuracy from the AI perspective and immutability from the blockchain perspective.

2.7. Validation Procedure

To ensure system reliability, a two-stage validation process was conducted: internal and external validation.

2.7.1. Internal Validation

Internal validation assessed the accuracy of AI-based halal classification by comparing the system's predictions against expert-labeled ground-truth data. Internal validation accuracy (V_{int}) was calculated as:

$$V_{int} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}(\hat{y}_i = y_i) \quad (17)$$

where N is the total number of samples, $\hat{y}_i$ is the predicted label, $y_i$is the expert label, and $\mathbb{I}(\cdot)$ is an indicator function that returns 1 when the prediction matches the expert label and 0 otherwise.

A high V_{int} score indicates strong agreement between the AI model and expert assessment.

2.7.2. External Validation

External validation assessed the transparency and integrity of the blockchain-based certificate mechanism. Each verification record was traced via QR code scanning to ensure public access to verification data and via Merkle root verification to ensure transaction integrity within a block. The Merkle root was verified using the formulation defined earlier in Eq. (11). When the computed Merkle root matched the stored Merkle root, the certificate record was considered valid and tamper-evident.

3. RESULTS

3.1. Experimental Setup

The prototype was implemented in Google Colab using Python 3.10. The core modules used include pytesseract (for OCR), transformers (for text classification), *opencv* (for image pre-processing), *hashlib* (for SHA-256 hashing), gradio (for the user interface), as well as reportlab and qrcode for creating PDF certificates and QR codes. A total of 214 cosmetic products from local and imported brands were analyzed, consisting of 142 halal (H), 38 haram (R), and 34 syubhat (S). Each record includes:

https://doi.org/10.31436/iiumej.v27i2.4059

- Product label image (JPEG/PNG format, 300 dpi resolution)
- List of extracted ingredients (manual input).

An example of the OCR-based ingredient extraction process is illustrated in Fig. 2, while the manual ingredient input interface is shown in Fig. 3.

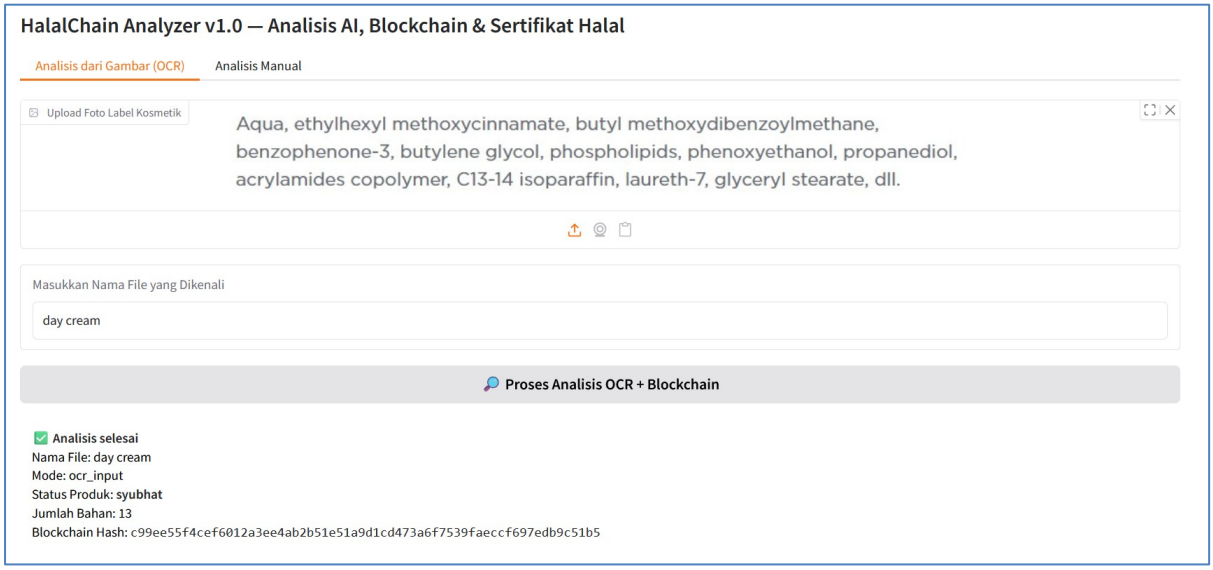


Figure 2. OCR Input

Ground-truth labels were established through expert review of ingredient lists, supported by authoritative halal ingredient databases and fiqh-based rulings on ingredients. Ingredients with ambiguous or disputed halal status were categorized as syubhat.

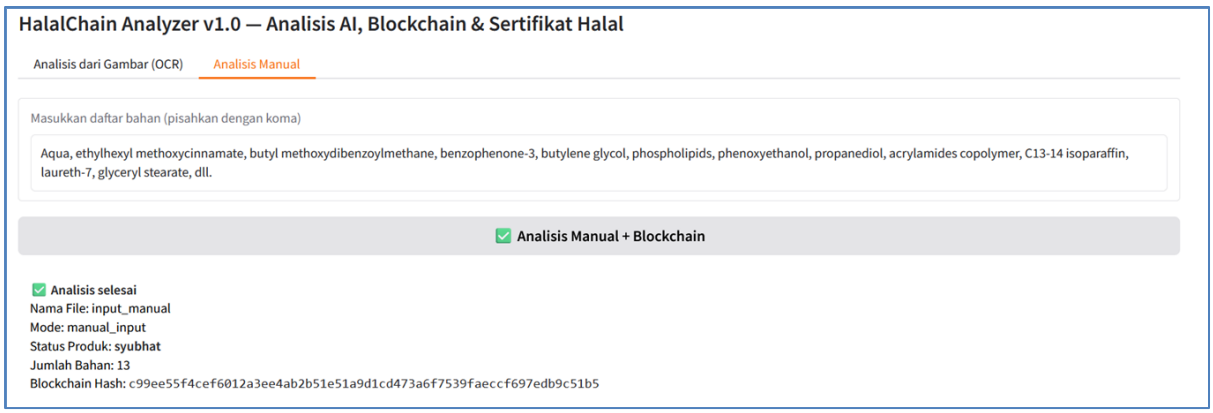


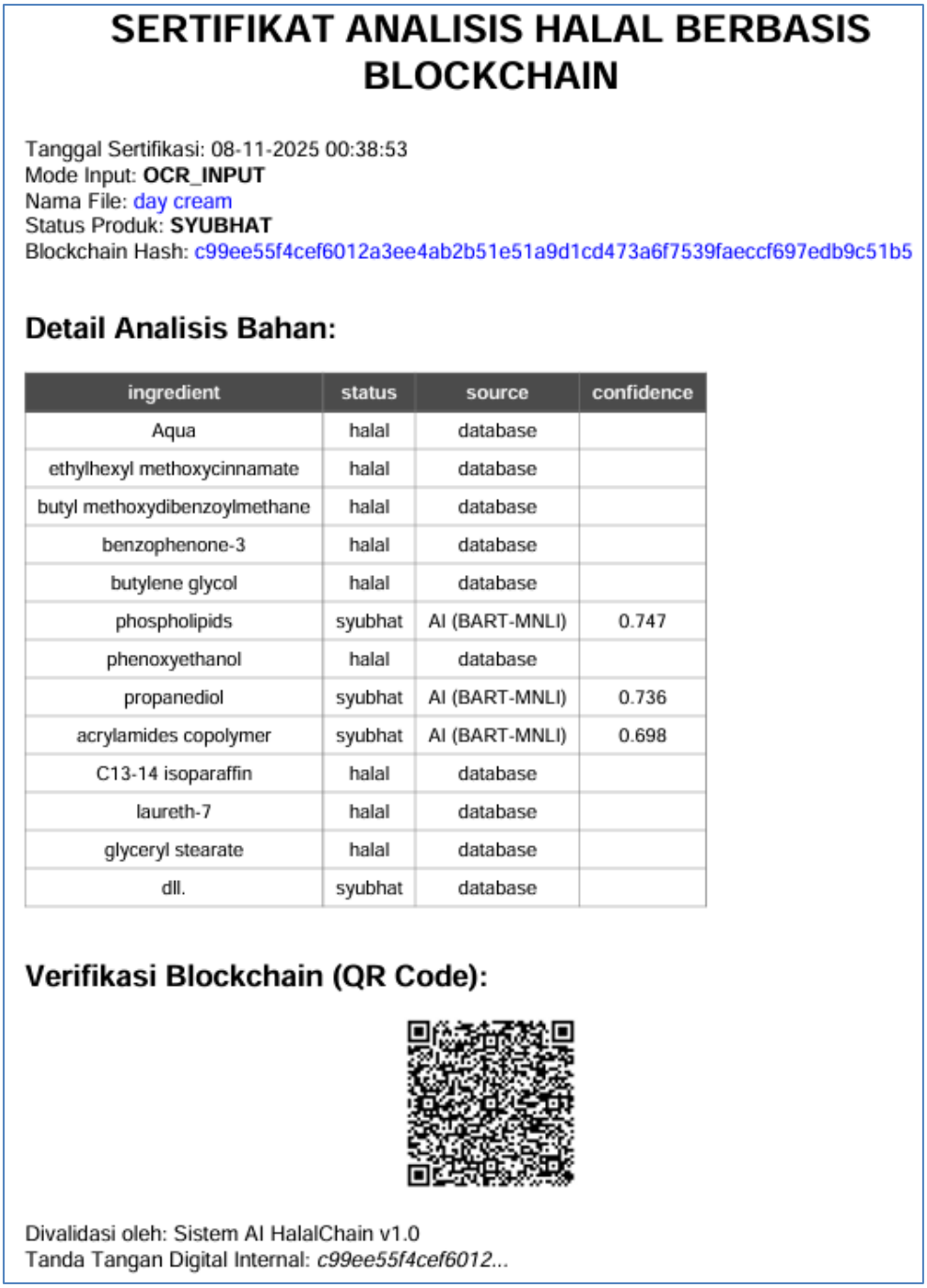
Figure 3. Manual Input

To evaluate the model, a 5-fold cross-validation technique was employed to ensure more consistent outcomes and preserve class balance across halal, haram, and syubhat. The dataset was split into training and testing portions for each validation stage, with no overlap. The average of all validation steps was used to record the final performance evaluator results.

The integrated OCR (Optical Character Recognition) pipeline achieves an average accuracy of 92.7% in recognizing text after several stages of pre-processing, including removing noise, identifying light and dark boundaries, and correcting letter shapes. Common OCR errors were observed under low lighting, densely packed or small fonts, and curved

https://doi.org/10.31436/iiumej.v27i2.4059

product surfaces. An example of the blockchain-based verification certificate generated by the system is shown in Fig. 4.



SERTIFIKAT ANALISIS HALAL BERBASIS BLOCKCHAIN

Tanggal Sertifikasi: 08-11-2025 00:38:53
Mode Input: **OCR_INPUT**
Nama File: day cream
Status Produk: **SYUBHAT**
Blockchain Hash: c99ee55f4cef6012a3ee4ab2b51e51a9d1cd473a6f7539faeccf697edb9c51b5

Detail Analisis Bahan:

ingredient	status	source	confidence
Aqua	halal	database	
ethylhexyl methoxycinnamate	halal	database	
butyl methoxydibenzoylmethane	halal	database	
benzophenone-3	halal	database	
butylene glycol	halal	database	
phospholipids	syubhat	AI (BART-MNLI)	0.747
phenoxyethanol	halal	database	
propanediol	syubhat	AI (BART-MNLI)	0.736
acrylamides copolymer	syubhat	AI (BART-MNLI)	0.698
C13-14 isoparaffin	halal	database	
laureth-7	halal	database	
glyceryl stearate	halal	database	
dll.	syubhat	database	

Verifikasi Blockchain (QR Code):

Divalidasi oleh: Sistem AI HalalChain v1.0
Tanda Tangan Digital Internal: *c99ee55f4cef6012...*

Figure 4. Certificate of Input OCR

3.2. Model Performance

The integrated Optical Character Recognition (OCR) pipeline achieved an average text-recognition accuracy of 92.7% after the pre-processing stage, which included noise removal, thresholding, and token normalization. Common recognition errors occurred due to low lighting, overly dense and small text, and curved product surfaces. The transformer-based text classification component achieved the performance results shown in Table 2. All reported metrics represent average performance across 5-fold cross-validation.

https://doi.org/10.31436/iiumej.v27i2.4059

Table 2. Average Performance Across 5-Fold Cross-Validation of the Text-Based Halal Classification Model

Metric	Value
Precision	90.2%
Recall	88.7%
F1-score	89.3%
Accuracy	91.2%

The verification process focuses primarily on ingredient composition extracted from OCR and manually entered text, since ingredient information is more reliable for halal verification than visual halal labels alone. In the proposed MXAI architecture, the text-based BERT component and image-based CNN component were combined using the weighted fusion mechanism defined earlier in Eq. (5). The fusion weight was set to $\alpha = 0.65$, indicating that the text modality contributed 65% to the final prediction, while the image modality contributed 35%. The value of $\alpha = 0.65$ was selected via grid search over values from 0.10 to 0.90 in increments of 0.05, as it yielded the best validation accuracy and maintained consistent SHAP explanations across cross-validation tests.

The proposed multimodal fusion model achieved an overall accuracy of 94.5%, outperforming both single-modal baseline models. The comparative results, averaged across 5-fold cross-validation, are presented in Table 3.

Table 3. Average Performance Across 5-Fold Cross-Validation: Accuracy Comparison Between Baseline Models and the Proposed MXAI+Blockchain System

Model Type	Accuracy	Improvement
Text-Only (BERT)	91,2%	-
Image-only (CNN)	87.6%	-
Proposed Fusion MXAI+Blockchain	94.5%	+3.3, improvement over the best baseline

The results show that the proposed multimodal approach surpasses the strongest single-modal baseline, namely the text-only BERT model, confirming the effectiveness of integrating textual and visual information within an MXAI and blockchain-based verification framework.

3.3. Comparison With Previous Research

To evaluate the novelty and effectiveness of the proposed Multimodal Explainable AI (MXAI) and Blockchain framework, a comparative analysis was conducted against prior research on halal verification and tracking. The comparison results are summarized in Table 4, highlighting differences in data modalities, explainability, blockchain integration, and reported accuracy across existing approaches and the proposed system.

Table 4. Comparison of Relevant Halal Verification and Tracking Methods

Method	Data	Explainability	Blockchain	Accuracy
Deep Learning [16]	Image data	no	no	91,8%
Comparative analysis [17]	Text and Image	no	no	72,5%
Evaluation [18]	Text and Image	yes	no	84%
Proposed MXAI+Blockchain System	Text dan Image	yes	yes	94,5%

Previous methods that focused on visual tracking and recognition, this research proposes integrating MXAI with Blockchain to ensure transparency and verifiable integrity. As shown

https://doi.org/10.31436/iiumej.v27i2.4059

in Table 4, existing studies primarily focus on processing single- or multimodal data without integrating explainability and blockchain-based integrity mechanisms. It is acknowledged that accuracy comparisons across studies may be influenced by variation in datasets and class distributions, particularly by the incorporation of explainable AI and blockchain-supported traceability, rather than asserting absolute performance superiority.

3.4. Blockchain Validation Results

The halal verification results were stored in a lightweight blockchain ledger using the SHA-256 hashing algorithm. Each block follows the structure defined earlier in Eq. (10), while hash consistency was validated using the formulation in Eq. (16). A total of 50 data records were verified in the simulation, with no inconsistencies detected. This indicates that all tested records maintained hash consistency during validation.

The Merkle tree structure was used for batch validation, as defined earlier in Eq. (11). This enables efficient verification of certificate records without exposing sensitive product data. All verified certificates generated a QR code linked to the blockchain hash and product verification record.

4. DISCUSSION

The research successfully developed a system prototype for verifying halal cosmetic products that integrates blockchain technology and Multimodal Explainable Artificial Intelligence (MXAI) to improve accuracy, transparency, integrity, and traceability. This prototype uses two data module sources: text from product labels and visual labels from OCR to produce halal, haram, or syubhat status classifications, achieving an accuracy of 94.5%.

This approach, which combines textual and visual labels of cosmetic products, improves classification results. Meanwhile, SHAP provides an explanation of AI concepts. Blockchain integration ensures that verification results are accurate, publicly accessible, and immutable.

4.1. Computational Cost and Scalability Considerations

From a computational perspective, the proposed Multimodal Explainable AI (MXAI) System is designed to balance accuracy and efficiency. The use of a pre-trained multilingual BERT-base model introduces moderate training overhead but remains efficient during inference due to the relatively short length of ingredient lists, while the image-processing stage is limited to OCR-based text extraction using a lightweight CNN. The multimodal fusion relies on simple weighted score aggregation, resulting in minimal additional computational cost. Training and evaluation were conducted in a cloud-based environment, and deployment-time complexity remains practical for large-scale use. The system can be scaled through batching or containerized microservices to support high-throughput halal verification.

4.2. Blockchain Implementation and Real-World Deployment Considerations

The blockchain component is implemented as a Python-based lightweight blockchain to demonstrate feasibility, transparency, and data integrity in a controlled setting. Only essential information, including halal status, document hash, timestamp, and Merkle root, is recorded to minimize storage overhead and latency. For real-world deployment, the framework can be extended to permissioned or consortium blockchain platforms where halal authorities and regulators act as validating nodes, enabling smart-contract-based certificate management and

auditing. By storing cryptographic hashes rather than raw data, the blockchain layer serves as a trust-and-integrity anchor without becoming a scalability bottleneck.

5. CONCLUSION

This research presents an integrated Multimodal Explainable Artificial Intelligence (MXAI) and blockchain framework for automated halal verification of cosmetic products. The proposed system combines ingredient information extracted through OCR, manually entered label text, and visual product-label features to classify products into halal, haram, or syubhat categories. The MXAI component improves decision transparency by providing explainable classification outputs, while the blockchain component strengthens traceability, integrity, and auditability through SHA-256 hashing, Merkle root verification, and QR-code-based certificate access. The experimental results show that the proposed fusion model achieved an accuracy of 94.5%, outperforming the text-only BERT and image-only CNN baselines. These findings indicate that integrating explainable AI with blockchain can improve trust in automated halal verification by making the classification process more accurate, transparent, and tamper-evident. The framework can be further extended to other halal supply chain sectors that require reliable product authentication, accountable decision-making, and secure digital certification.

REFERENCES

[1] A. Azwar and Abur Hamdi Usman, "Artificial Intelligence In The Halal Industry: Trends And Global Research Opportunities," Int. J. Halal Ind., vol. 1, no. 1, pp. 46–67, 2025, doi: 10.20885/ijhi.vol1.iss1.art4.

[2] A. Masood, S. R. H. Hati, and A. A. Rahim, "Halal Cosmetics Industry for Sustainable Development: a Systematic Literature Review," Int. J. Bus. Soc., vol. 24, no. 1, pp. 141–163, 2023, doi: 10.33736/ijbs.5609.2023.

[3] F. Omar, A. Nabot, and R. Alqirem, "Transforming Decision Support Systems Through Artificial Intelligence: Enhancing Analytics, Automation, and Interaction," 2025, pp. 479–491. doi: 10.1007/978-3-031-90271-0_34.

[4] A. K. B. Arnob, R. R. Chowdhury, N. A. Chaiti, S. Saha, and A. Roy, "A Comprehensive Systematic Review Of Intrusion Detection Systems: Emerging Techniques, Challenges, and Future Research Directions," J. Edge Comput., vol. 4, no. 1, pp. 73–104, 2025, doi: 10.55056/jec.885.

[5] N. Nurhayati, A. Chelbi, A. Ben Mabrouk, and F. C. Young, "Unleashing Halal Potential: Exploring The Dynamics of Halal Logistics In Tangerang City, Indonesia," Agrociencia, pp. 1–13, Mar. 2025, doi: 10.47163/agrociencia.v59i3.3272.

[6] F. Sunmola, G. Baryannis, A. Tan, K. Co, and E. Papadakis, "Holistic Framework for Blockchain-Based Halal Compliance in Supply Chains Enabled by Artificial Intelligence," Systems, vol. 13, no. 1, pp. 1–26, 2025, doi: 10.3390/systems13010021.

[7] K. Kumar, V. Kumar, Seema, M. K. Sharma, A. A. Khan, and M. J. Idrisi, "A Systematic Review of Blockchain Technology Assisted with Artificial Intelligence Technology for Networks and Communication Systems," J. Comput. Networks Commun., vol. 2024, pp. 1–15, Feb. 2024, doi: 10.1155/2024/9979371.

[8] V. T. Hoang et al., "Halal or Not: Knowledge Graph Completion for Predicting Cultural Appropriateness of Daily Products," IEEE Access, vol. 13, no. December 2024, pp. 15158–15167, 2025, doi: 10.1109/ACCESS.2025.3528488.

[9] A. Ghosh et al., "Blockchain-Assisted Serverless Framework for AI-Driven Healthcare Applications," Int. J. Adv. Comput. Sci. Appl., vol. 16, no. 5, pp. 473–482, 2025, doi: 10.14569/IJACSA.2025.0160546.

[10] S. Bhatia and A. S. Albarrak, "A Blockchain-Driven Food Supply Chain Management Using QR Code and XAI-Faster RCNN Architecture," Sustain., vol. 15, no. 3, 2023, doi: 10.3390/su15032579.

[11] R. Salama, D. Cacciagrano, and F. Al-Turjman, "Connecting AI and Blockchain to Improve Security of Financial Services," Barolli, L. Adv. P2P, Parallel, Grid, Cloud Internet Comput., vol. 232., doi: https://doi.org/10.1007/978-3-031-76462-2_7.

[12] M. Vasileiou et al., "Digital Transformation of Food Supply Chain Management Using Blockchain: A Systematic Literature Review Towards Food Safety and Traceability," Bus. Inf. Syst. Eng., 2025, doi: 10.1007/s12599-025-00948-0.

[13] Nor Aina Mardhiah Nazim and Syarifah Md Yusof, "Halal Legislation Enforcement of Food Premises in Malaysia," Glob. J. al thaqafah, vol. SI, no. 1, pp. 151–164, Jul. 2023, doi: 10.7187/GJATSI072023-12.

[14] S. Wicaksono, M. K. H. Al-Asy'ari, E. D. N. El-Adibah, and Iffahasanah, "MAQASHID SHARIA PROGRESSIVE: Anatomical and Transformational of Halal Institutions in UIN KHAS Jember," El-Mashlahah, vol. 13, no. 2, pp. 107–132, 2023, doi: 10.23971/el-mashlahah.v13i2.7370.

[15] D. K. Elwaked, N. M. Kassim, and R. Thurasamy, "Enhancing Internal Integration Through AI: A Resource-Based View With Data-Driven Culture As Moderator," J. Logist. Informatics Serv. Sci., vol. 12, no. 2, pp. 209–228, 2025, doi: 10.33168/JLISS.2025.0213.

[16] N. Zulkeefli, Siti Nur Izzati, Hashim, "Comparison of CNN-based Algorithms for Halal Logo Recognition," J. Syst. Manag. Sci., vol. 12, no. 5, pp. 155–168, 2022, doi: 10.33168/JSMS.2022.0510.

[17] S. Al-Thani, Shaikha Nasser;Anjum, "Comparative performance of ChatGPT, Gemini, and final-year emergency medicine clerkship students in answering multiple-choice questions: implications for the use of AI in medical education," Int. J. Emerg. Med., vol. 18, no. 1, p. 146, 2025, doi: 10.1186/s12245-025-00949-6.

[18] M. T. Zamir, S. U. Khan, and S. U. Khan, "Explainable AI-Driven Analysis of Radiology Reports Using Text and Image Data: Experimental Study," JMIR Form. Res., vol. 9, 2025, doi: 10.2196/77482.

https://doi.org/10.31436/iiumej.v27i2.4061

Generative AI Models: A Comparison of Application Analysis on Web AI-Based Decision Support Systems for Satellite Anomaly Identification

ABDUL MUTHOLIB[1,2], NADIRAH ABDUL RAHIM[1*], TEDDY SURYA GUNAWAN[1], AHMAD SHAH HIZAM MD YASIR[3]

[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*
[3]*Faculty of Resilience, Rabdan Academy, Abu Dhabi, United Arab Emirates*

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

(Received: 24 November 2025; Accepted: 19 January 2026; Published online: 10 May 2026)

ABSTRACT: The rapid innovation of Generative Artificial Intelligence (GenAI) has transformed Decision Support Systems (DSS) across various domains, including satellite operations. This paper presents a comparative analysis of four free Generative AI models, including Gemma 3 by Google, Llama 4 Maverick by Meta AI, Nemotron Nano 2 by NVIDIA, and Devstral Small by Mistral, in the context of integrating them into a Web AI-based Decision Support System (DSS) for satellite anomaly identification. Using a dataset of over 4,455 satellite anomaly records covering 1957 to 2024 provided by Seradata, with scalability and adaptability to diverse mission profiles. We evaluate these models by generating anomaly analyses for clarity, accuracy, completeness, and relevance across the Incident Overview, Reliability Trend, Insight, and Stakeholder Recommendation categories, using a 5-point Likert scale and Fleiss' Kappa for internal consistency. The comprehensive evaluation of GenAI models for Web AI-based DSS delineates a clear performance stratification, with scores of 4.44 and 4.39 for Nemotron Nano 2 and Llama 4 Maverick, respectively, confirming their positions as the leading systems based on overall Likert scores. However, the analysis further revealed a critical trade-off between absolute quality and internal consistency (Fleiss' Kappa). The superior models, Nemotron Nano 2 and Llama 4 Maverick, achieved high Likert scores by displaying pronounced performance peaks but suffered the lowest internal predictability, with $\kappa = 0.18$ on Llama 4 Maverick, indicating a highly volatile output structure in which strong clarity often masked critical incompleteness. Conversely, Devstrall Small, despite its suboptimal mean score of 3.44, demonstrated the highest internal consistency, with $\kappa = 0.66$. This robust predictability, even at a lower level, underscores a significant implication for DSS development. The model selection must prioritize the required balance between absolute performance ceiling and the predictability of the output structure. The findings highlight the potential of GenAI implementation in DSS to enhance the reliability of satellite operations, while exploring future directions for research and development in this area. This research contributes to the development of more resilient and intelligent satellite anomaly identification systems, with broader implications for space mission safety, resource optimization, cost reduction, and the future of AI-driven aerospace technologies.

ABSTRAK: Inovasi pesat Kecerdasan Buatan Generatif (GenAI) telah mengubah Sistem Sokongan Keputusan (DSS) merentasi pelbagai domain, termasuk operasi satelit. Kajian ini membentangkan analisis perbandingan empat model AI Generatif percuma termasuk: Gemma 3 oleh Google, Llama 4 Maverick oleh Meta AI, Nemotron Nano 2 oleh NVIDIA dan Devstral

Small oleh Mistral, dalam menyepadu Sistem Sokongan Keputusan (DSS) berasaskan web AI bagi mengenal pasti anomali satelit. Mengguna pakai lebih 4,455 rekod set data anomali satelit meliputi tahun 1957 hingga 2024 oleh Seradata, set data ini mempunyai ciri kebolehskalaan dan kebolehsuaian pada profil misi yang pelbagai. Kajian ini menilai kesemua model melalui penjanaan analisis anomali melalui kejelasan, ketepatan, kelengkapan dan kerelevanan merentasi kategori Gambaran Keseluruhan Insiden, Trend Kebolehpercayaan bagi, Wawasan dan Cadangan Pihak Berkepentingan, mengguna skala Likert 5-mata dan Kappa Fleiss kekonsistensi dalaman. Penilaian komprehensif model GenAI DSS berasaskan web AI menggariskan stratifikasi prestasi yang jelas pada skor 4.44 dan 4.39 untuk Nemotron Nano 2 and Llama 4 Maverick masing-masing, mengesahkan keupayaan sebagai sistem utama berdasarkan skor Likert keseluruhan. Walau bagaimanapun, analisis selanjutnya mendedahkan pertukaran kritikal antara kualiti mutlak dan kekonsistensi dalaman (Fleiss' Kappa). Model terbaik, Nemotron Nano 2 and Maverick Llama 4, mencapai skor Likert tertinggi dengan puncak prestasi yang tinggi tetapi keboleh ramalan dalaman terendah iaitu $\kappa \approx 0.18$ pada Maverick Llama 4, menandakan dapatan struktur yang sangat tidak menentu di mana kejelasan yang kuat sering menutupi ketidaklengkapan kritikal. Sebaliknya, Mistral Devstrall Small menunjukkan kekonsistensi dalaman tertinggi $\kappa \approx 0.66$ walaupun skor min suboptimumnya 3.44. Keboleh ramalan yang mantap ini, walaupun pada tahap lebih rendah, menggariskan implikasi ketara bagi pembangunan DSS. Pemilihan model mesti mengutamakan keseimbangan yang diperlukan antara siling prestasi mutlak dan keboleh ramalan dapatan struktur. Penemuan ini mengetengah potensi pelaksanaan GenAI DSS dalam meningkatkan kebolehpercayaan operasi satelit sambil meneroka hala tuju masa depan bagi tujuan penyelidikan dan pembangunan. Penyelidikan ini menyumbang kepada pembangunan sistem mengenal pasti anomali satelit berdaya tahan dan pintar, dengan implikasi yang lebih luas bagi keselamatan misi angkasa lepas, mengoptimum sumber, mengurangkan kos operasi dan masa depan teknologi aero angkasa yang dipacu AI.

KEYWORDS: *Generative AI, Gemma 3, Llama 4 Maverick, Nemotron Nano 2, Devstral Small, Satellite Anomaly Identification, Clarity, Accuracy, Completeness, Relevance, Decision Support Systems.*

1. INTRODUCTION

The escalating complexity of satellite operations demands robust, efficient anomaly-identification systems. Satellites, essential for communication, navigation, Earth observation, and scientific research, produce substantial quantities of telemetry data. Manually combing through this vast amount of data to identify anomalies from normal behavior is not only time-consuming but also prone to human mistakes. Conventional anomaly identification techniques often struggle with the complexity, high dimensionality, and class imbalance of the satellite dataset, which was labor-intensive and prone to human error [1]. This can lead to delayed identification, resulting in significant operational downtime, costly repairs, or even mission loss.

The growing field of Artificial Intelligence (AI), particularly Generative AI (GenAI) and Large Language Models (LLMs), offers a promising paradigm shift in addressing these challenges. These advanced models possess an unparalleled ability to learn intricate patterns, understand contextual nuances, and even generate human-readable explanations from complex datasets [1].

Web-based decision support systems (DSS) further enhance the advantage of generative models by enabling real-time collaboration, visualization, and automated analysis. However, the integration of lightweight, state-of-the-art generative models, such as Gemma 3, Llama 4 Maverick, Nemotron Nano 2, and Devstral Small, into web-based DSS for satellite anomaly

https://doi.org/10.31436/iiumej.v27i2.4061

identification remains largely unexplored. While these models are designed for efficiency and performance, their comparative effectiveness in high-stakes, real-time environments, particularly in terms of clarity, accuracy, completeness, and relevance, has not been systematically evaluated.

Despite the rapid advancement of GenAI, its application, especially concerning DSS for satellite anomaly identification, faces several challenges:

- *Clarity*: Models must provide interpretable outputs to support human validation, yet few studies quantify interpretability across different architectures.
- *Accuracy*: High precision and recall are critical, but comparative analyses of these models' performance on satellite anomaly data are lacking.
- *Completeness*: Models must cover a wide range of anomaly types, but their ability to do so has not been systematically assessed.
- *Relevance*: Generated insights must be contextually appropriate for operational decision-making, a dimension rarely addressed in existing literature.

Prior research has primarily focused on larger models, such as GPT, Gemini, Grok, and DeepSeek, and not focused on real applications such as DSS, especially in the satellite anomaly area [2], leaving a gap in understanding how smaller, efficient generative models perform in constrained, real-time environments. The incorporation of GenAI into Decision Support Systems (DSS) is particularly promising for satellite anomaly identification, an essential task for assuring mission success and spacecraft longevity. Such a system could empower satellite stakeholders not only with anomaly identification but also with insightful, AI-generated diagnoses, significantly enhancing their decision-making capabilities.

This study addresses these gaps through a quantitative comparison of Gemma 3, Llama 4 Maverick, Nemotron Nano 2, and Devstral Small in a web-based DSS for satellite anomaly identification and evaluates the implementation of certain GenAIs within the DSS, identifying the recommended GenAI models for DSS for satellite anomaly identification.

The remainder of this paper is organized as follows: Section 2 describes the GenAI models, and Section 3 presents the research methodology for integrating these GenAI models into the Web AI-based DSS. Section 4 illustrates a comparative analysis of how GenAI models are implemented in our Web AI-based DSS. Finally, Section 5 concludes the findings, discusses the challenges and future GenAI models for DSS applications, and suggests directions for future research.

2. GENERATIVE AI IN DSS FOR SATELLITE ANOMALY IDENTIFICATION

2.1. Satellite Anomaly Identification

Satellite operations are inherently complex, involving numerous interconnected subsystems that must operate seamlessly in the hostile space environment [3, 4]. Satellite anomalies denote any deviation from the expected or nominal behavior of a satellite's components or overall system [3]. These can range from subtle sensor drifts to catastrophic failures in critical systems such as power generation, attitude control, or communication. Common categories found in the Seradata database related to satellite anomalies can be summarized as follows [3]:

- *Power System*: Fluctuations in solar array output, battery degradation, or power distribution issues.

https://doi.org/10.31436/iiumej.v27i2.4061

- *Attitude Control System*: Unintended rotations, thruster malfunctions, or gyroscope failures affecting satellite orientation.
- *Thermal Control System*: Overheating or undercooling of components.
- *Communication System*: Signal degradation, transponder failures, or antenna pointing errors.
- *On-board Data / Computer*: Software glitches, memory errors, or processor malfunctions.

Traditional methods for anomaly identification have many approaches, including statistical process control (SPC), trend analysis, and rule-based expert systems [5, 6]. While these methods are straightforward, they often suffer from high false-positive rates due to normal operational variations or miss true anomalies that manifest as subtle, non-linear patterns [5, 6]. They also require extensive domain expertise to define rules and thresholds, making them time-consuming and difficult to scale.

2.2. Generative AI (GenAI)

Satellite operations are inherently complex, involving numerous interconnected subsystems that GenAI refers to a class of artificial intelligence models capable of producing novel content, such as text, images, or data, that resembles the distribution of the data they were trained on. GenAI took on a specific form following the publication of Google's research in 2017 [7]. A team of scientists developed a neural network known as the Transformer that performed exceptionally in language translation, and carefully designed it to take advantage of the capabilities of parallel computing technology known as graphics processing units (GPUs) [7]. Large Language Models (LLMs) are a prominent subset of GenAI, characterized by their vast number of parameters and training on massive text datasets, enabling them to understand, generate, and manipulate human language with remarkable fluency and coherence. Beyond text generation, LLMs have demonstrated capabilities in complex pattern recognition, reasoning, summarization, and explanation generation, making them highly relevant for sophisticated data analysis tasks like anomaly identification. GenAI systems were initially single-modal, anticipating the input to possess a particular modality and generating output of a designated modality. In this case, the most popular GenAI systems were text-in, text-out single-modal systems and text-in, image-out single-modal systems [7].

GenAI comprises the hardware and software required to develop, train, and deploy models. Graphics Processing Unit, known as GPU, is a key part of the hardware supporting GenAI, with Nvidia as the principal entity [7]. The extensive matrix multiplication activities and the logic layers fundamental to neural networks are ideally suited to Nvidia's GPU architecture. In addition, the CUDA Software Development Kit (SDK) supports fast parallel processing by directly accessing sets of instructions in the GPU hardware [7].

This study focuses on a comparative analysis of four free GenAI models, including Gemma, Llama Maverick, Nemotron, and Devstral Small. The potential of these models for satellite anomaly identification lies not only in their ability to detect deviations but also in their capacity to interpret complex data patterns, synthesize information, and generate actionable explanations for operators. This moves beyond simple alerts to providing “why” an anomaly is occurring, which is crucial for effective decision-making.

2.2.1. Google Gemma 3

Gemma 3 is a family of advanced multimodal models developed by Google DeepMind, co-designed with the Gemini frontier model family [8]. Gemma 3 models are comparable in size to Gemma 2, with an additional 1B-parameter model and numerous new capabilities, including multimodality, long context, and multilingual [8]. In the context of Web AI-based

https://doi.org/10.31436/iiumej.v27i2.4061

DSS, multimodality enables simultaneous correlation of text data with a visual chart, while long context enables ingestion of the entire provided historical dataset. Gemma 3 comes in two size options: 7-billion- and 2-billion-parameter models. A 7 billion parameter model suitable for efficient development and deployment on GPU and TPU, while a 2 billion parameter model customized for CPU and on-device applications. Both parameter sizes are designed to adopt distinctive computational limitations, applications, and developer requirements [9].

The Gemma 3 models are constructed following the Transformer decoder architecture [10], as is illustrated in Fig. 1. Gemma 3 does not need the encoder part of the original Transformer architecture; it needs only the decoder, which is equivalent to an encoder, except for MASKING in the multi-head attention block, where the decoder only allows information to be collected from the previous words in the sentence [10]. Gemma 3 refined through structural enhancements and model optimizations to manage a 128K context window and training on up to 6T tokens [8]. To handle this scale, Gemma 3 employs Grouped-Query Attention (GQA), which decreases memory operating cost by sharing data across multiple processing heads, and Root Mean Square Layer Normalization (RMSNorm) in both pre-norm and post-norm configurations [8]. Pre-norm applies normalization before the attention or feed-forward layer, leading to significantly more stable training at enormous scales, while post-norm applies it afterward the remaining connection, which can improve the model's final representational power [8]. The models are powerful and adaptable, significantly outperforming their predecessors [8]. These architectural refinements are essential for Decision Support Systems (DSS) and anomaly identification, as they provide the structural stability and memory efficiency.

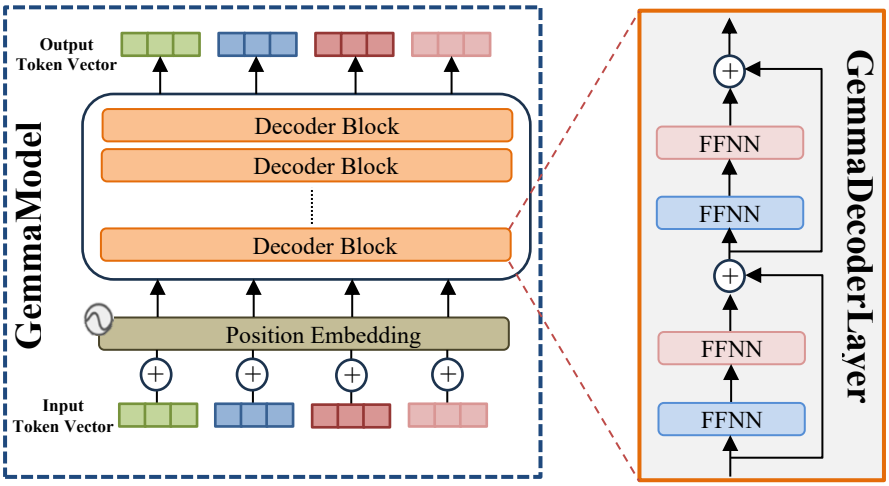


Figure 1. Gemma 3 model architecture.

2.2.2. Meta AI Llama 4 Maverick

Llama 4 Maverick is a high-capacity multimodal language model from Meta with 128K tokens and 17 billion active parameters [11]. It supports multilingual text and image input, and produces multilingual text and code output across 200 supported languages [11]. Building upon the success of previous Llama iterations, Llama 4 Maverick is expected to feature significant advancements in scale, efficiency, and reasoning capabilities.

Llama 4 Maverick is constructed using a mixture-of-experts (MoE) architecture as depicted in Fig. 2 [11]. In this architecture, a single token activates only a portion of the total

https://doi.org/10.31436/iiumej.v27i2.4061

parameters and exhibits superior computational efficiency for both training and inference, delivering higher quality compared to a dense model [11]. A Router analyses incoming data tokens and dynamically assigns them to specific Routed Experts or subnetworks that have been trained to specialize in niche areas such as mathematical reasoning or linguistic nuances [11]. Llama 4 Maverick employs a combination of dense and mixture-of-experts (MoE) layers to improve interpretability. The MoE comprises 128 routed experts in addition to a shared expert, with each token concurrently processed by the shared expert and one routed expert [11]. While the complete parameter set resides in memory, only a subset is activated during inference. As a result, this improves inference efficiency by reducing computational processing costs and latency [11]. This setup enables the model to sustain a large-scale knowledge base while dynamically engaging only the most appropriate experts for a given query. Such efficiency is particularly advantageous in real-time Decision Support Systems (DSS) and anomaly detection tasks, where rapid response and high precision are critical for identifying rare data anomalies.

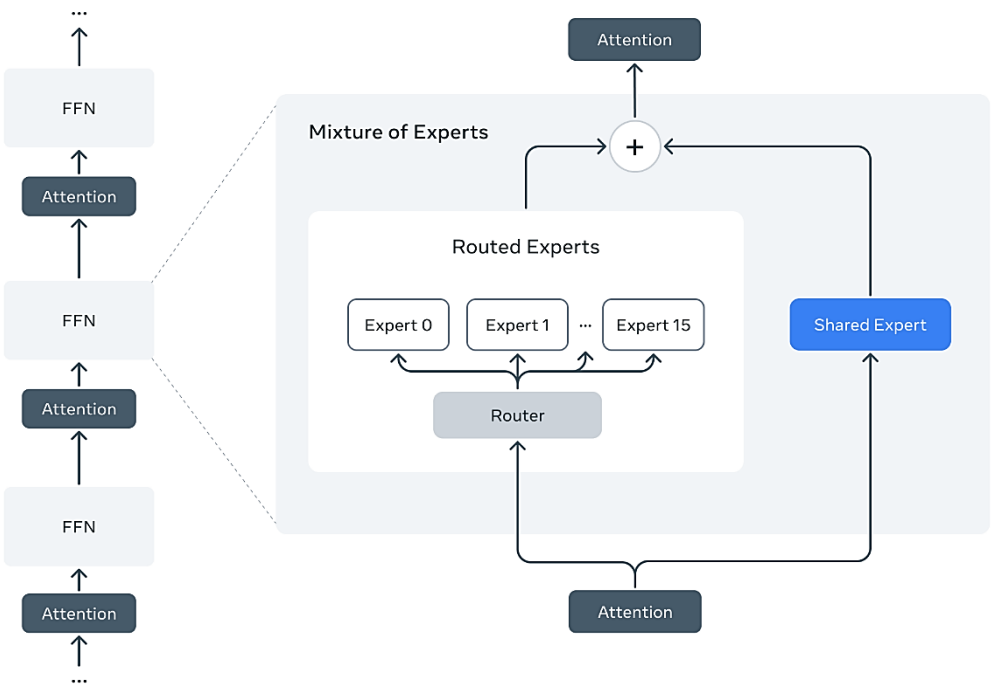


Figure 2. Llama 4 Maverick Mixture of Experts (MoE) architecture

2.2.3. NVIDIA Nemotron Nano 2

Nemotron Nano 2 is a diverse, open collection of thinking models created by NVIDIA, offering outstanding reasoning capabilities, efficient inference, and an open license for commercial applications [12]. Nemotron Nano 2 is a hybrid Mamba-Transformer reasoning model developed by NVIDIA to improve benchmark accuracy and handle heavy scenarios such as 1k input, 8k output, 8k input, and 16k output tokens [12]. It builds based on the architecture of Nemotron-H with new datasets and recipes for pre-training, aligning, pruning, and distillation [12].

The illustration of the Nemotron Nano 2 base-layer pattern, as shown in Fig. 3, comprises Mamba 2, self-attention, and FFN layers [12]. The architectural design comprises three distinct phases, with a total of 62 layers: 6 attention layers, 28 FFN layers, and 28 Mamba-2 layers [12]. The initial phase includes three repetitions of a Mamba-2 and feed-forward network (FFN) pair. This is followed by a central core, reiterated six times, which incorporates a single

https://doi.org/10.31436/iiumej.v27i2.4061

self-attention layer followed by a combination of Mamba-2 and FFN layers. The final phase concludes with a single Mamba-2 and FFN pair. Notably, the design applies only six attention layers, which is about 8% of the total depth, while relying on 28 Mamba-2 layers for linear-time processing. This selective distribution enables the model to deliver up to a sixfold higher implication throughput compared to standard dense architectures, while preserving competitive accuracy on complex reasoning tasks.

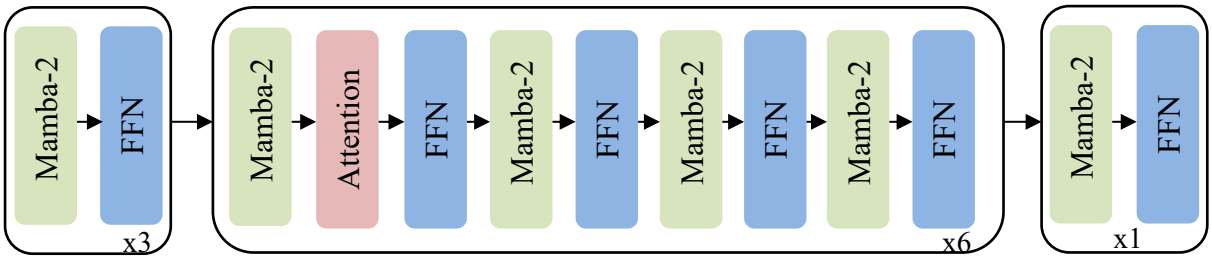


Figure 3. Nemotron Nano 2 Base Layer Pattern

Nemotron Nano 2 was pre-trained using DeepSeek's FP8 training recipe, which has precision over 20T tokens using a Warmup-Stable-Decay (WSD) learning rate schedule, and then undertook an endless pre-training long context addition phase to become 128K capable deprived without degrading other standard [12]. The post-training of Nemotron Nano 2 utilized a combination of Supervised Fine-Tuning (SFT), Group Relative Policy Optimization (GRPO), Direct Preference Optimization (DPO), and Reinforcement Learning from Human Feedback (RLHF), and was performed on roughly 90B tokens [12].

2.2.4. Mistral Devstral Small

Devstral is an agentic Large Language Model (LLM) for software engineering tasks, built in collaboration between Mistral AI and All Hands AI. Devstral excels at using tools to explore codebases, editing multiple files, and powering software engineering agents [13]. Devstral Small, as a more compact model, having only 24B parameters, could offer a compelling balance between performance and computational resource requirements [13]. Its efficiency could be particularly advantageous for deployment in web-based systems where resource optimization is a key consideration, while still retaining strong analytical capabilities for anomaly identification.

Devstral Small is a compact yet powerful Transformer model with 24 billion parameters. Built based on Mistral Small 3, it features 40 layers and incorporates Grouped Query Attention (GQA) for efficient processing [13]. The model is trained on an extensive range of textual data, including both natural language and programming code. Its capabilities are further enhanced through a long context extension phase, enabling it to handle up to 128,000 tokens, which is ideal for tasks involving code agents and other applications demanding extended context windows [13] such as a Decision Support System (DSS).

The Devstral Small model is evaluated using the OpenHands scaffold, an open platform designed to develop AI agents that interact with the world similarly to human software developers [13]. It offers a comprehensive framework that allows agents to write code, interact with command lines, and browse the web within a secure, sandboxed environment.

2.3. Web AI-Based Decision Support Systems

A Decision Support System (DSS) is an information system that aids decision-making. It typically comprises three main components: a data management subsystem for data collection and organization, a model management subsystem for analytical tools and models, and a user interface subsystem for interaction [14]. Web-based DSS leverage internet technologies to provide accessible, scalable, and collaborative decision-making environments [15]. Their advantages include:

- *Accessibility*: Users can access the system from anywhere with an internet connection.
- *Scalability*: Easily accommodates a growing number of users and data.
- *Collaboration*: Facilitates shared access to information and collaborative decision-making.
- *Cost-effectiveness*: Reduces the need for specialized client-side software.

The role of AI in enhancing DSS capabilities is profound. AI can transform a traditional DSS into an intelligent DSS by automating data analysis, providing predictive insights, offering prescriptive recommendations, and generating explanations and descriptions [4, 16]. For satellite anomaly identification, an AI-based DSS can process vast streams of telemetry data, identify potential issues, and present them to operators with clear, concise explanations, thereby augmenting human cognitive abilities and accelerating response times. While specific public examples of Web AI-based DSS exclusively for satellite anomaly identification using advanced GenAI are limited, similar systems exist in other domains like industrial predictive maintenance, energy, healthcare, manufacturing, and satellite, demonstrating the feasibility and benefits of such an integration [17-20].

2.4. Systematic Literature Mapping

The uniqueness of this study, compared to previous studies presented in Table 1, lies in its focused comparative analysis of GenAI models (Gemma 3, Llama 4, Maverick, Nemotron Nano 2, and Devstral Small) integrated with a Web AI-based DSS designed for satellite anomaly identification. This research provides an empirical, multi-model evaluation tailored to the unique challenges and requirements of satellite operations, emphasizing not only identification accuracy but also the crucial aspect of explainability provided by GenAI models. This comprehensive approach aims to provide valuable insights into which GenAI models are best suited to enhance the resilience and efficiency of DSS for future satellite missions.

Table 1. Systematic Literature Mapping

Title	Author(s)	Contribution	Methods	Gap
Grok, Gemini, ChatGPT, and DeepSeek: Comparison and Applications in Conversational Artificial Intelligence	Murillo Edson de Carvalho Souza et al. [2]	Comparative analysis of LLMs (Grok, Gemini, ChatGPT, DeepSeek)	Transformer-based LLMs	Lack of standardized evaluation metrics across models; limited transparency in proprietary systems
Generative AI as a New Platform for Application Development	Michael A. Cusumano, Annabelle Gawer, David B. Yoffie [7]	GenAI as a foundational platform	Ecosystem mapping, governance	Insufficient frameworks for sustainable governance and open-source integration

Prototype Development of Web-AI-Based Decision Support System: Insights and Recommendations for Satellite Anomaly Identification	A. Mutholib, N. A. Rahim, and T. S. Gunawan [16]	Web-based DSS prototype for anomaly identification in satellites	Web-AI integration, anomaly identification models	Prototype stage, lacks large-scale validation
Generative AI in AI-Based Digital Twins for Fault Diagnosis for Predictive Maintenance in Industry 4.0/5.0	E. Mikołajewska, D. Mikołajewski, T. Mikołajczyk, and T. Paczkowski [17]	Digital twins for predictive maintenance	GANs, VAEs, Transformers	Limited real-world deployment evidence, need for federated and edge-based implementations
Using Machine Learning for Advanced Anomaly Detection and Classification	Lane et al. [19]	Anomaly detection in SSA	SVM, RF, k-NN, PCA	Lack of generative approaches for anomaly simulation; limited DSS integration for operational decision-making
Preliminary Design and Methodological Framework for an AI-Driven Decision Support System in Earth Observation Satellites	Abdallah Alabed, Tarık Özkul [18]	Framework for onboard DSS in EO satellites; drone-based simulation for cloud-aware imaging	YOLOv12 (cloud detection), SORT (tracking), adaptive mission planner	Limited to simulation; lacks orbital validation; planner rule-based (no reinforcement learning yet)

3. METHODOLOGY

This section discusses the methodical approach adopted to develop the Web AI-based Decision Support System (DSS) for satellite anomaly identification and to conduct the comparative analysis of the selected generative AI models.

3.1. System Architecture of the Web AI-Based DSS

The implementation of Web AI-based DSS uses a three-tier architecture, as illustrated in Fig. 4, including the frontend as the presentation layer, the backend as the logic tier, and the database as the data tier.



Figure 4. Web AI-based DSS Three-tier Architecture.

The frontend, known as the User Interface (UI), is built with modern web technologies, such as the Bootstrap framework, to provide an intuitive, interactive dashboard for satellite stakeholders. The UI displays anomaly data in both graph and table formats. It includes features for reviewing historical data, applying filters, and generating GenAI-generated analysis.

https://doi.org/10.31436/iiumej.v27i2.4061

The backend consists of an API, and the logic is implemented using a robust framework: Python with Django, to handle data ingestion, process frontend requests, orchestrate interactions with GenAI models, and manage the database. A PostgreSQL database is used to store satellite data, anomaly data, and GenAI-generated analysis.

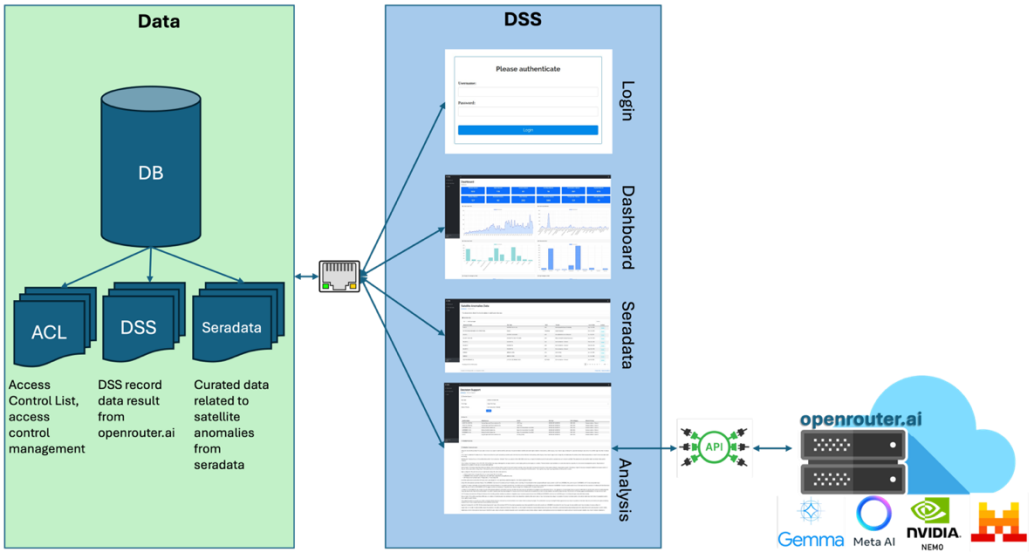


Figure 5. Web AI-based DSS System Architecture.

In addition, the system's high-level structure is defined by a blueprint that outlines how its components will interact and work together to meet the requirements. The details of the system architecture are illustrated in Fig. 5. The figure shows three sections: data, DSS, and Openrouter, which serve as GenAI models API providers.

The data section comprises a database that includes access control management, DSS analysis, and curated data from Seradata. This data is accessible via the local area network. It means that the data and the DSS application are in the same area network. This local access includes the authentication process, dashboard access, data presentation, and part of the data analysis. The other part of the data analysis requires an external connection or internet access.

The analysis of data for decision-making uses the API provided by openrouter.ai to access GenAI models, including Gemma 3, Meta AI Llama 4 Maverick, Nvidia Nemotron Nano 2, and Mistral Devstral Small.

3.2. Dataset Selection

The effectiveness of any AI-driven anomaly identification system is heavily dependent on the quality and representativeness of its data. The data source used in this research is the Seradata SpaceTrak service [20]. Seradata is an open-source intelligence database that contains extensive information regarding international satellite and space launch vehicle activities, including failure reporting [3]. The selection of Seradata for this research is necessary due to its extensive depth, historical breadth, and thorough failure reporting, which are crucial for a meticulous analysis of satellite reliability [4]. With more than six decades of recorded data, Seradata provides a comprehensive perspective on satellite performance and failure characteristics, enabling researchers to understand the immediate and long-term effects of design decisions, operating conditions, and external environmental factors on satellite reliability [20].

https://doi.org/10.31436/iiumej.v27i2.4061

Spanning from 1957 to 2024, Seradata provides over 4,455 occurrences of spacecraft failures, rendering it an essential resource for examining trends and patterns in satellite anomaly identification [4]. The data used in this study have been preprocessed to minimize missing values, noise, and the possibility of outliers by using data cleansing. This preprocessing is very important to enhance raw data quality, consistency, and efficiency for further processing [3, 4]. The dataset classifies failures into nine main categories: attitude control, power, payload instrument, beam, control processor, telemetry, thermal, and transponder, providing a solid basis for analyzing the causes of satellite faults. Table 2 describes the satellite anomaly datasets used in this study.

Table 2. Description of satellite anomaly data

Attributes	Values
Number of failure categories	9
Data duration	1957 - 2024
Number of experimental datasets	4455

3.3. Generative AI Model Integration and Customization

Integrating various GenAI models into a DSS requires careful evaluation of their APIs and optimal prompting techniques. In this study, OpenRouter is utilized for model invocations, employing the following specific models as presented in Table 3, and all configured with default settings:

Table 3. GenAI models specification

Model Name	Params	Context length	Knowledge cutoff	Source
Gemma 3 [21]	27B	1M	December 2024	`google/gemma-3-27b-it:free`
Llama 4 Maverick [22]	17B	1M	August 2024	`meta-llama/llama-4-maverick-17b-128e-instruct:free`
Nemotron Nano 2 [23]	253B	131K	April 2025	`nvidia/nemotron-nano-9b-v2:free`
Devstral Small 2505 [24]	24B	33K	May 2025	`mistralai/mistral-small-3.2-24b-instruct:free`

Given the nature of LLMs, effective prompt engineering was essential to guide each model toward accurate anomaly identification and meaningful explanation generation. Prompts were carefully designed to encompass:

- *Contextual Information*: Current anomaly readings, historical trends, and relevant satellite operational modes.
- *Task Definition*: Clear instructions to identify anomalies and provide a concise explanation of the anomaly, its potential cause, its impact, and recommendations.

Each GenAI model was integrated into the Web AI-based DSS backend via its corresponding RESTful API. The responses from models' API, typically JSON objects containing the generated text, were subsequently parsed and processed by the Explanation Generation Module.

https://doi.org/10.31436/iiumej.v27i2.4061

3.4. Experimental Setup and Evaluation Metrics

The evaluation of each GenAI model was conducted in a controlled simulation environment to ensure reproducible, comparable results.

3.4.1. Simulation Environment

The Web AI-based DSS was deployed on both local and cloud platforms for the GenAI models API, with sufficient computational resources and GPUs on the local platform to accelerate inference and handle the workload of querying multiple LLMs simultaneously. The simulation involved feeding the preprocessed satellite anomaly data through the DSS. This setup allowed for accurate measurement of resource utilization under operational load.

Table 4. Hardware Specifications

Category	Specification
CPU	Intel(R) Core i9-14900K, 3.20 GHz
GPU	NVIDIA GeForce RTX 4090
RAM	64.0 GB DDR5
Storage (SSD)	500 GB
Storage (HDD)	6 TB

Table 4 presents the hardware configuration details, featuring cutting-edge components optimized for demanding computational operations. The Intel Core i9-14900K CPU, with a clock speed of 3.20 GHz, provides substantial processing power and efficiency for data preprocessing and model training. The NVIDIA GeForce RTX 4090 GPU enhances parallel processing, crucial for intricate ML algorithms and deep learning models. The machine, equipped with 64 GB of DDR5 RAM, accommodates large-scale memory-intensive tasks, such as processing the massive Seradata dataset. Storage comprises a 500 GB SSD for high-speed access to commonly utilized files and a 6 TB HDD for long-term storage, facilitating efficient data management.

The software environment presented in Table 5 was designed to meet the computational requirements of the Web AI-based DSS, ensuring speed, efficiency, and reproducibility. Python 3.11.9 as the foundation, Django 5.1.3 as the back-end web framework, PostgreSQL 14 as the database management system (DBMS), and Bootstrap 5 as the front-end toolkit are used to construct a Web AI-based DSS.

Table 5. Software Specifications

Category	Version
Windows Server Datacenter	2022
Python	3.11.9
Django	5.1.3
PostgreSQL	14
Bootstraps	5

3.4.2. Evaluation Metric

A comprehensive quantitative metric was used to evaluate each GenAI model's performance. A Likert scale and Fleiss' Kappa (κ) were employed in this study to provide a more comprehensive explanation of model performance. A Likert scale was utilized for scoring metrics, which provides five possible answers, ranging from 1 (strongly disagree) to 5 (strongly agree), to gauge the responses from the expert users who participate in the questionnaire to

indicate positive or negative strength of agreement of feeling regarding the question or statement [25].

Fleiss' Kappa is a statistical measure used to assess the level of agreement between three or more users who rate or respond when they classify or score items using categorical scales, such as a Likert scale with 1 to 5 categories [26]. To assess the robustness of the expert evaluation, Fleiss' Kappa was computed to quantify agreement among the expert evaluators. Fleiss' Kappa (κ) is suitable for assessing consistency when multiple raters assign categorical ratings. Fleiss' Kappa (κ) is formulated as in the following equation 1 [26].

$$k = \frac{\bar{P}_a - \bar{P}_c}{1 - \bar{P}_c} \quad (1)$$

where $\bar{P}_a$ is the mean agreement between raters and $\bar{P}_c$ is the mean probability of agreement on change. The metric quantifies agreement beyond chance, with commonly accepted thresholds of $\kappa \geq 0.41$ indicating moderate agreement, $\kappa \geq 0.61$ substantial agreement, and $\kappa \geq 0.81$ almost perfect agreement [26].

Both the Likert scale and Fleiss' Kappa were crafted to ensure comprehensive coverage of the dimensional metrics as presented in Table 6.

Table 6. Quantitative Metrics Description

Metrics	Description
Clarity	The explanation is easy to understand.
Accuracy	The explanation accurately describes the anomaly and its cause, or has signs of hallucination.
Completeness	The explanation provides sufficient detail for decision-making.
Relevance	The explanation is relevant to the observed anomaly data.

3.5. Questionnaire Responders

The expert users selected for this study typically represent an interdisciplinary cohort of spacecraft engineers, mission controllers, and space scientists with 5 to 10 years' experience and a proven ability to perform cross-domain diagnostics. Their nature is described by a deep experiential understanding of satellite health. The experts have backgrounds in satellite anomalies, particularly in spacecraft subsystems such as Attitude Control Systems (ACS), Electrical Power Systems (EPS), and communication systems, and have spent years studying signal anomalies that may indicate sensor noise, environmental interference such as solar flares, or genuine hardware degradation. Their primary role in the questionnaire process is to provide the qualitative weights and diagnostic logic that bridge the gap between raw statistical outliers and actionable operational decisions.

4. Results and Discussions

This section presents the empirical findings from the comparative analysis of Gemma 3, Llama 4 Maverick, Nemotron Nano 2, and Devstral Small within the developed Web AI-based Decision Support System (DSS) for satellite anomaly identification. We discuss each model's individual performance, conduct a direct comparison across key metrics, provide insights from the DSS implementation, and acknowledge the study's inherent limitations. The evaluation was conducted using a questionnaire and the Web AI-based DSS by a human expert in the satellite field.

https://doi.org/10.31436/iiumej.v27i2.4061

4.1. Performance of Generative AI Models

Assessing the performance of GenAI models is more complex than that of conventional classification models. In a generative model, there is no single "correct" answer; instead, there is a continuum of acceptable outputs. However, the core concepts of precision, recall, and F1 score can be adapted to evaluate different aspects of a generative model's performance [27]. The evaluation disclosed distinct performance profiles for each generative AI model across various anomaly types and operational scenarios within the simulated satellite environment.

In this study, several scenarios were applied to prompts and GenAI models that were implemented in the Web AI-based DSS. Table 7 presents the scenarios applied.

Table 7. Prompts Scenarios

Scenarios	Prompts
Incident overview	As an expert in the satellite area, write the incident overview related to the mission's primary, orbit category, mass at launch, age, and design life, especially for ages less than design life, based on the data provided.
Reliability Trends & Insights	As an expert in the satellite area, write the Reliability Trends & Insights related to mission primary, orbit category, mass at launch, age, and design life, especially for ages less than design life, based on the data provided.
Stakeholder Recommendations	As an expert in the satellite area, write the Stakeholder Recommendations related to mission primary, orbit category, mass at launch, age, and design life, especially for ages less than design life, based on the data provided.

4.1.1. Likert Scale Analysis

Fig. 5 shows the overall model performance across all categories (Incident Overview, Reliability Trend, Insights, and Stakeholder Recommendations) and metrics (Clarity, Accuracy, Completeness, and Relevance).

From Fig. 6, Nemotron Nano 2 outperforms other models, scoring 4.4 out of 5, followed closely by Llama 4 Maverick, with an overall score of 4.39 out of 5. It shows that both models consistently scored between "Agree" and "Strongly Agree". On the other hand, Gemma 3 shows moderate performance, while Devstral Small achieved the lowest score among the models, averaging closer to "Neutral" and "Agree."

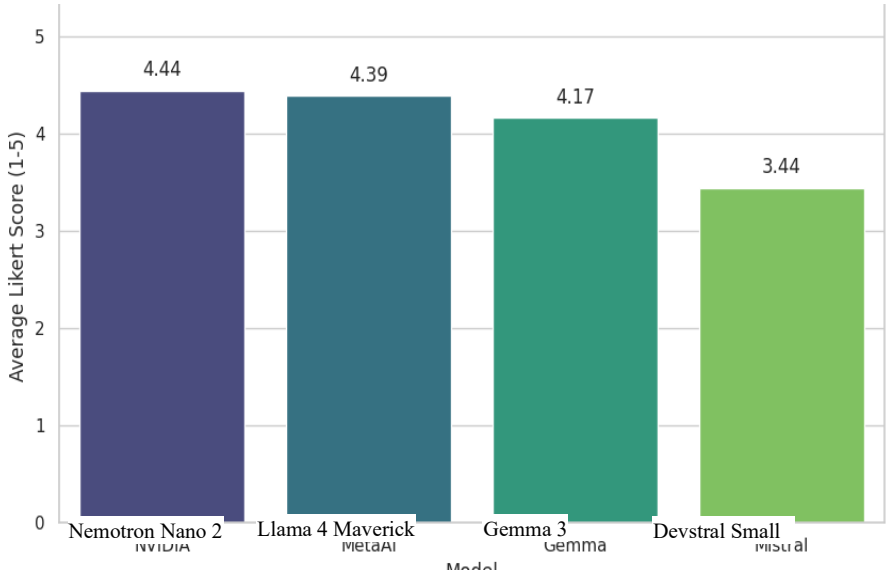


Figure 6. Overall Average Assessment Score by Model

https://doi.org/10.31436/iiumej.v27i2.4061

The categorical assessment, as shown in Fig.7, reveals a distinct hierarchy in model efficacy, with Nemotron Nano 2 and Llama 4 Maverick consistently outperforming their counterparts across the evaluated domains of Incident, Reliability, and Recommendation.

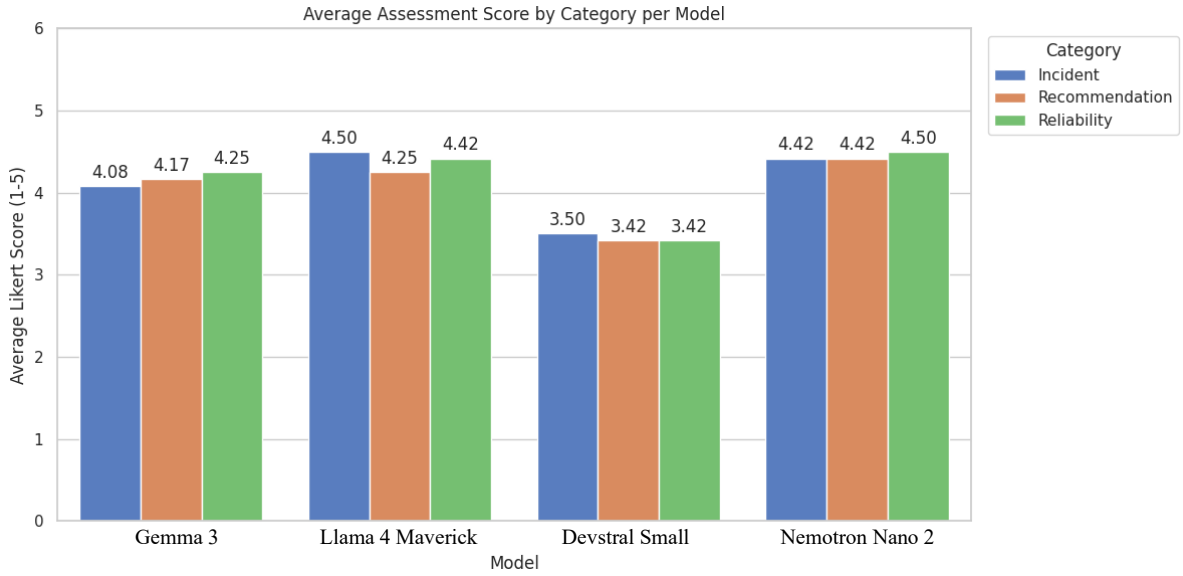


Figure 7. Average Assessment Score by Category per Model

Nemotron Nano 2 established pre-eminence in the Reliability and Recommendation categories, achieving mean Likert scores of 4.50 and 4.42, respectively, indicating a robust capacity for consistent and actionable output. Conversely, Llama 4 Maverick demonstrated a slight advantage in the Incident category, achieving 4.50, suggesting specialized proficiency in handling event-specific queries. While Gemma 3 maintained competitive performance thresholds, particularly in Reliability, Devstrall Small exhibited a marked divergence from the leading models, recording the lowest aggregate scores across all categories, notably falling to 3.42 in both Reliability and Recommendation assessments.

The heatmap for the detailed average assessment score by metric and category is presented in Fig. 8. It can be summarized that the heatmap presents a granular comparative analysis of model performance across the distinct dimensions of Incident, Reliability, and Recommendation, further ranked by the qualitative metrics of Accuracy, Clarity, Completeness, and Relevance. Visualization uses a continuous color gradient to represent the mean Likert-scale evaluations, with darker hues corresponding to higher performance ratings. This matrix shows that Nemotron Nano 2 and Llama 4 Maverick consistently achieve high scores, especially in the domains of Reliability and Incident overview, as evidenced by a predominance of values above 4.33. In contrast, the lighter color associated with Devstral Small across multiple vectors is most pronounced in Recommendation Accuracy and Completeness, revealing a significant performance disparity and underscoring its suboptimal assessment relative to the benchmark established by the leading models.

https://doi.org/10.31436/iiumej.v27i2.4061



Figure 8. Detailed Average Assessment Score by Metric and Category

4.1.2. Inter-Rater Reliability Results

The inter-rater reliability analysis yielded a Fleiss' Kappa for Gemma 3 as presented in Table 8, showing that Gemma 3 demonstrates substantial internal consistency for the Reliability (κ=0.692) and Recommendation (κ=0.636) dimensions. This means that when the model performs well, it tends to be correspondingly strong across Clarity, Accuracy, Completeness, and Relevance. However, the notably lower score for the Incident dimension (κ=0.404) highlights a specific performance imbalance during complex tasks. The data shows that Gemma 3 often maintains high scores for Clarity and Relevance, declaring confidence and on topic, even when its Accuracy and Completeness drop slightly. This suggests that while Gemma 3 is a communicative and reliable tool overall, its hallucination risk is highest during specific incidents where it prioritizes sounding clear over being perfectly accurate.

Table 8. Fleiss' Kappa Analysis for Gemma 3

Dimension	Fleiss' Kappa (κ)	Agreement Level	Interpretation
Incident	**0.404**	Moderate	Significant consistency break.
Reliability	**0.692**	Substantial	Highly consistent performance.
Recommendation	**0.636**	Substantial	Strong consistency.

Based on Table 9, Llama 4 Maverick exhibits low internal consistency on complex tasks, as evidenced by the slight-to-fair agreement scores for Incident (κ=0.180) and Reliability (κ=0.219). While the Recommendation score (κ=0.692) remains strong, the detailed metrics

https://doi.org/10.31436/iiumej.v27i2.4061

expose a critical gap in challenging scenarios; the model maintains perfect scores for Clarity and Relevance but suffers a significant drop in Completeness. This statistical divergence points to a superficial fluency issue: Llama 4 Maverick generates highly sophisticated, on-topic responses that sound convincing but fail to capture all necessary information, resulting in answers that are clear but materially incomplete.

Table 9. Fleiss' Kappa Analysis for Llama 4 Maverick

Dimension	Fleiss' Kappa (κ)	Agreement Level	Interpretation
Incident	**0.180**	Slight	Extreme divergence between metrics.
Reliability	**0.219**	Fair	Very low consistency.
Recommendation	**0.692**	Substantial	Strong consistency.

In Table 10, Nemotron Nano 2 demonstrates Fair internal consistency (κ≈ = 0.32–0.38), which is noticeably lower than that of the previous models. The Fair agreement here is not due to random noise; it is due to a consistent performance gap in complex tasks. This score is driven by a systematic Form over Substance split in challenging scenarios, where the model maintains good scores for Clarity and Relevance, but drops to average scores for Accuracy and Completeness. Unlike Gemma 3, which had high highs or Llama 4 Maverick, which had polished omissions, Nemotron Nano 2's data suggests a general degradation of performance in complex tasks, where both the delivery Clarity and the content Accuracy suffer, but the content suffers more.

Table 10. Fleiss' Kappa Analysis for Nemotron Nano 2

Dimension	Fleiss' Kappa (κ)	Agreement Level	Interpretation
Incident	**0.317**	Fair	Moderate split between form and substance.
Reliability	**0.385**	Fair	Weak consistency, but systematic.
Recommendation	**0.317**	Fair	Consistency mirrors Incident performance.

Devstral Small exhibits the highest consistency, reaching Substantial agreement for Incident and Recommendation as shown in Table 11. This is because the scores across all four criteria are very tightly grouped. This high Kappa score indicates the model is highly predictable; if it is accurate, it is also clear, complete, and relevant. The caveat is that its absolute performance ceiling is lower than the other models, but its performance is balanced across all four metrics.

Table 11. Fleiss' Kappa Analysis for Devstral Small

Dimension	Fleiss' Kappa (κ)	Agreement Level	Interpretation
Incident	**0.657**	Substantial	Highest consistency for Incident metric.
Reliability	**0.556**	Moderate	Strong overall consistency.
Recommendation	**0.657**	Substantial	Highly consistent.

To sum up, the comprehensive evaluation of GenAI models for Web AI-based DSS reveals a critical trade-off between absolute quality and internal consistency. Nemotron Nano 2 and Llama 4 Maverick emerge as the superior systems in terms of absolute quality, achieving overall Likert scores of 4.44 and 4.39, respectively. However, this superior performance comes at the cost of output predictability, as evidenced by their low Fleiss' Kappa scores (κ≈ 0.32 for Nemotron Nano 2 and κ≈ 0.18 for Llama 4 Maverick). This low consistency suggests that, while these models achieve high scores, the reliability of their output structure is volatile, with

significant internal discrepancies, such as the gap between Llama 4 Maverick's Clarity (Strongly Agree) and Completeness (Neutral). In contrast, Devstral Small, which recorded a suboptimal Likert score of 3.44, indicating a need for substantial improvement in absolute quality, demonstrated the highest internal consistency ($\kappa = 0.66$). This suggests that while Devstral Small's output quality is lower, it is highly predictable, with all criteria rising and falling in a uniform manner. Gemma 3 performed as a competitive mid-tier option, scoring 4.17 and exhibiting moderate consistency ($\kappa = 0.40$), balancing strong performance with moderate internal predictability. Ultimately, model selection must therefore prioritize the required balance between absolute performance ceiling (Likert Score) and output predictability (Fleiss' Kappa) for the specific DSS application.

4.2. Insights from Web AI-Based DSS Implementation

The development and deployment of the Web AI-based DSS delivered valuable insights for integrating GenAI models into operational decision-making workflows. One important challenge was optimizing prompt engineering to consistently obtain accurate anomaly identifications and relevant explanations from each LLM. Different models responded best to varying prompt structures and levels of detail, demanding iterative refinement. Another challenge involved managing API request limits and ensuring robust error handling for continuous data streams. Real-time data processing needs careful orchestration to minimize latency between telemetry ingestion and anomaly presentation to the user.

Despite these challenges, the implementation was a significant success in representing the feasibility and benefits of such a system. The web-based interface proved highly accessible and intuitive for simulated stakeholders, including owner, manufacturer, and operator. The integration of GenAI explanations alongside anomaly data visualizations significantly improved stakeholder awareness and reduced the time required to understand anomaly nature.

The feedback from the experts was overwhelmingly positive. Users highlighted the clarity and helpfulness of the GenAI explanations, the AI-based DSS significantly reduced diagnostic time, and provided a starting point for investigation that was not available before. The AI-based DSS effectively transformed raw anomaly data into actionable intelligence, demonstrating its potential to enhance operational efficiency and resilience.

4.3. Limitations of the Study

While this study provides valuable insights, it is important to acknowledge certain limitations:

- *Simulated Data*: The study primarily utilized a simulated satellite anomaly dataset. While designed to mimic real-world conditions and anomaly types, it may not fully capture the full complexity, noise, and unforeseen events present in live satellite data. Future work will aim to validate these findings in real-world settings, where permissible.
- *Model Versions*: The performance of GenAI models is continuously evolving. The results presented are based on the specific versions of Gemma 3, Llama 4 Maverick, Nemotron Nano 2, and Devstral Small available at the time of this research. Newer iterations may exhibit different performance characteristics.
- *Scope of Anomaly Types*: While a diverse set of anomaly types was injected, the study did not cover every conceivable satellite malfunction. The models' performance on highly novel or infrequent anomalies might vary.

- *Prompt Engineering Dependency*: The performance of LLMs is highly sensitive to prompt engineering. Although significant effort was invested in optimizing prompts, different prompting strategies may yield slightly different results.
- The Fleiss' Kappa score indicates only fair agreement among evaluators. This may be due to the subjective nature of assessing generative text explanations. Future work will expand the evaluator pool and refine scoring guidelines to reduce variability in interpretation.
- *Computational Resources*: The resource utilization metrics are specific to the cloud environment and configurations used in this study. Performance may vary in different hardware or infrastructure setups.

These limitations highlight areas for future research and development to further enhance GenAI's robustness and applicability for satellite anomaly identification.

5. CONCLUSIONS AND FUTURE WORK

This paper presented a comprehensive comparative analysis of four state-of-the-art GenAI models, Gemma 3, Llama 4 Maverick, Nemotron Nano 2, and Devstral Small, within a Web-based AI-based Decision Support System (DSS) for satellite anomaly identification. The study successfully demonstrated the feasibility and significant benefits of integrating advanced GenAI into this critical domain, moving beyond simple identification to providing actionable, AI-generated explanations.

The empirical findings reveal a clear performance hierarchy in model efficacy. Nemotron Nano 2 and Llama 4 Maverick emerged as the superior architectures, demonstrating high-fidelity performance across all evaluated metrics, with Nemotron Nano 2 establishing a benchmark in the Reliability Trend, Insight, and Stakeholder Recommendation tasks (avg. > 4.40). The data suggests that while current top-tier models are highly capable of handling complex incident overview and reliability queries, significant performance disparities remain among standard open-weight models. Notably, Devstral Small's consistent underperformance (avg. 3.44), particularly in the accuracy and completeness of recommendations, underscores the critical need for rigorous pre-deployment validation.

Crucially, Fleiss' Kappa analysis of internal consistency presents a necessary counterpoint to the absolute performance scores. The highest-scoring models, Llama 4 Maverick $\kappa \approx 0.18$, Slight Agreement, and Nemotron Nano 2 $\kappa \approx 0.32$, Fair Agreement, exhibit the lowest output predictability. This indicates that their superior Likert scores are often achieved at the cost of internal structural reliability, marked by significant discrepancies between criteria, such as high Clarity masking low Completeness. In direct contrast, Devstral Small, despite its consistent underperformance (avg. 3.44), demonstrated the highest output consistency ($\kappa = 0.66$), Substantial Agreement. This suggests Devstral Small's, though suboptimal in quality, is the most predictable model, with all output metrics aligning tightly.

The primary implication of this research is that selecting a GenAI model for a Web AI-based DSS in satellite operations is not a one-size-fits-all decision. The optimal model depends directly on the mission's priorities: for critical, high-value assets where detailed diagnostics are essential. This study fills a significant research gap by providing the first empirical evidence of these performance trade-offs, thereby guiding future development and deployment strategies for intelligent satellite anomaly-identification systems.

This research lays a foundation for further advancements in AI-driven satellite operations. Future work should focus on several key areas, including prioritizing the expansion of the

evaluation corpus to include a wider diversity of incident scenarios to verify whether Nemotron Nano 2's reliability advantage holds across edge cases. Additionally, investigating the specific failure modes of the Devstal Small model through qualitative error analysis could provide insights into architectural limitations. Further research could also explore fine-tuning strategies or Retrieval-Augmented Generation (RAG) implementations to potentially bridge the performance gap observed in mid-tier models like Gemma 3 and Devstal Small.

ACKNOWLEDGEMENT

This research is fully supported by the Asian Office of Aerospace Research and Development (AOARD) under the grant scheme numbers FA2386-23-1-4073 and SPI23-179-0179.

REFERENCES

[1] N. Kazanskiy, R. Khabibullin, A. Nikonorov, and S. Khonina, "A Comprehensive Review of Remote Sensing and Artificial Intelligence Integration: Advances, Applications, and Challenges," Sensors, vol. 25, no. 19, p. 5965, 2025. [Online]. Available: https://www.mdpi.com/1424-8220/25/19/5965.

[2] M. Edson de Carvalho Souza and L. Weigang, "Grok, Gemini, ChatGPT and DeepSeek: Comparison and Applications in Conversational Artificial Intelligence," vol. 1, 02/18 2025, doi: 10.5281/zenodo.14885243.

[3] A. Mutholib, N. A. Rahim, T. S. Gunawan, and A. A. Ahmarofi, "Performance Comparison of Data Preprocessing Methods for Trade-Space Exploration with AI Model: Case Study of Satellite Anomalies Detection," in 2024 IEEE 10th International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA), 2024: IEEE, pp. 271-275.

[4] A. Mutholib, N. A. Rahim, T. S. Gunawan, and M. Kartiwi, "Trade-Space Exploration with Data Preprocessing and Machine Learning for Satellite Anomalies Reliability Classification," IEEE Access, 2025.

[5] P. H. Tran, A. Ahmadi Nadi, T. H. Nguyen, K. D. Tran, and K. P. Tran, "Application of machine learning in statistical process control charts: A survey and perspective," in Control charts and machine learning for anomaly detection in manufacturing: Springer, 2022, pp. 7-42.

[6] H. Akbarian, "Deep Learning Based Anomaly Detection in Space Systems and Operations," Florida Atlantic University, 2024.

[7] Cusumano, M. A., V. F. Farias, and R. Ramakrishnan, "Generative AI as a New Platform for Applications Development," An MIT Exploration of Generative AI no. September, 2024, doi: https://doi.org/10.21428/e4baedd9.f189351f.

[8] G. Team et al., "Gemma 3 technical report," arXiv preprint arXiv:2503.19786, 2025.

[9] G. Team et al., "Gemma: Open models based on gemini research and technology," arXiv preprint arXiv:2403.08295, 2024.

[10] T. Kao. "Why Gemma adopts the Decoder-Only Transformer architecture?" https://makerpro.cc/2024/04/why-gemma-adopts-decoder-only-transformer-architecture/ (accessed 20 October, 2025).

[11] "The Llama 4 herd: The beginning of a new era of natively multimodal AI innovation." https://ai.meta.com/blog/llama-4-multimodal-intelligence/ (accessed 11/07/2025.

[12] A. Basant et al., "Nvidia nemotron nano 2: An accurate and efficient hybrid mamba-transformer reasoning model," arXiv preprint arXiv:2508.14444, 2025.

[13] A. Rastogi et al., "Devstral: Fine-tuning Language Models for Coding Agent Applications," arXiv preprint arXiv:2509.25193, 2025.

[14] E. Turban, J. E. Aronson, and T. P. Liang, Decision Support Systems and Intelligent Systems, 7th Edition ed. New Delhi: Prentice Hall of India, 2007.

[15] R. Islam et al., "The future of cloud computing: benefits and challenges," International Journal

[16] A. Mutholib, N. A. Rahim, and T. S. Gunawan, "Prototype Development of Web AI-Based Decision Support System: Insights and Recommendations for Satellite Anomaly Identification," in Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS), Maui, Hawaii, September 16-19 2025.

[17] E. Mikołajewska, D. Mikołajewski, T. Mikołajczyk, and T. Paczkowski, "Generative AI in AI-Based Digital Twins for Fault Diagnosis for Predictive Maintenance in Industry 4.0/5.0," Applied Sciences, vol. 15, no. 6, p. 3166, 2025. [Online]. Available: https://www.mdpi.com/2076-3417/15/6/3166.

[18] A. Alabed and T. Özkul, "Preliminary Design and Methodological Framework for an AI-Driven Decision Support System in Earth Observation Satellites," in 2025 7th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (ICHORA),

[19] B. Lane, M. Poole, M. Camp, and J. Murray-Krezan, "Using machine learning for advanced anomaly detection and classification," in Advanced Maui Optical and Space Surveillance Tech. Conf.(AMOS), 2016.

[20] T. M. Grile and R. A. Bettinger, "Statistical reliability estimation for satellites operating from 1991-2020 with payload reliability focus," in 2022 6th International Conference on System Reliability and Safety (ICSRS), 2022: IEEE, pp. 378-386.

[21] "Google: Gemma 3 27B (free)." https://openrouter.ai/google/gemma-3-27b-it:free (accessed 16 October, 2025).

[22] "Llama 4 Maverick 17B 128E Instruct." https://huggingface.co/meta-llama/Llama-4-Maverick-17B-128E-Instruct (accessed 11/07/2025.

[23] N. Corporation. "NVIDIA-Nemotron-Nano-9B-v2." https://huggingface.co/nvidia/NVIDIA-Nemotron-Nano-9B-v2 (accessed 12 October, 2025).

[24] J. Denize. "Devstral Small" https://huggingface.co/mistralai/Devstral-Small-2505 (accessed 11/07/2025.

[25] A. Marengo, F. G. Karaoglan-Yilmaz, R. Yılmaz, and M. Ceylan, "Development and validation of generative artificial intelligence attitude scale for students," Frontiers in Computer Science, vol. 7, p. 1528455, 2025.

[26] R. Cole, "Inter-Rater Reliability Methods in Qualitative Case Study Research," Sociological Methods & Research, vol. 53, no. 4, pp. 1944-1975, 2024, doi: 10.1177/00491241231156971.

[27] A. Bonnet. "AI Metrics that Matter: A Guide to Assessing Generative AI Quality." https://encord.com/blog/generative-ai-metrics/ (accessed 15/09, 2025).

https://doi.org/10.31436/iiumej.v27i2.4124

PromptLessSAM: From Foundational Model to Domain Expert via Lightweight Decoder Adaptation for Crack Segmentation

MOHAMMED AL-MUSTAFA*, ISRAA H. ALI

Software Department, College of Information Technology, University of Babylon, Babylon, Iraq

**Corresponding author: mohammedalmustafa.hendo@student.uobabylon.edu.iq*

(Received: 10 January 2026; Accepted: 20 April 2026; Published online: 10 May 2026)

ABSTRACT: Maintaining infrastructure such as roads and bridges is very important for safety. This is called Structural Health Monitoring (SHM). A key part of SHM is automatically finding cracks in pavements. Recent foundation models, such as the Segment Anything Model (SAM), can segment a wide range of object types. However, they are not well-suited for specific tasks such as detecting thin cracks in complex environments. They also require user input to function, making them unsuitable for fully automatic inspection systems. Therefore, this paper proposes PromptLessSAM to transform the general SAM model into a domain expert for crack segmentation. We use a lightweight method in which we freeze the powerful image encoder of SAM, except for the encoder's neck, which remains trainable, and we add a new, small decoder. Our model is very efficient and has only about 1.1 million trainable parameters. We trained and tested our model on the public Crack500 dataset. The experimental results show that our model achieved high performance (a Dice score of 72.7% and an IoU of 57.11% on Crack500), outperforming many state-of-the-art models, including TransUNet and PaveSAM. Also, our model showed good generalization capabilities on the DeepCrack dataset (Dice score of 74.7% and IoU of 62.43%), which it did not see during training. PromptLessSAM provides a new, efficient, and effective method. It shows how to adapt large foundation models for a specific task. Our work provides a strong, high-performance solution trained on a very small number of parameters.

ABSTRAK: Penyelenggaraan infrastruktur seperti jalan raya dan jambatan sangat penting bagi tujuan keselamatan. Ini dikenali sebagai Pemantauan Kesihatan Struktur (Structural Health Monitoring, SHM). Salah satu komponen utama pada SHM ialah mengesan rekahan pada turapan secara automatik. Model asas terkini, seperti Model Sesuatu Bahagian (SAM), boleh melakukan segmentasi pelbagai objek. Namun, model ini kurang berkesan bagi tugas khusus seperti mengesan rekahan yang nipis pada persekitaran kompleks. Selain itu, SAM memerlukan arahan daripada pengguna untuk berfungsi, menjadikannya kurang sesuai bagi sistem pemeriksaan sepenuhnya automatik. Oleh itu, kajian ini mencadangkan PromptLessSAM untuk menukarkan model SAM yang umum kepada model pakar domain bagi segmentasi rekahan. Kajian ini menggunakan kaedah yang ringan dengan membekukan pengekod imej SAM yang berkuasa, kecuali pada bahagian 'leher' pengekod yang kekal dilatih, dan menambah sebuah penyahkod baharu yang kecil. Model ini sangat cekap dan hanya mempunyai kira-kira 1.1 juta parameter boleh latih. Model ini dilatih dan diuji menggunakan set data awam Crack500. Keputusan eksperimen menunjukkan model ini mencapai skor prestasi tinggi (iaitu skor Dice sebanyak 72.7% dan IoU sebanyak 57.11% pada set data Crack500), mengatasi banyak model tercanggih (state-of-the-art) seperti TransUNet dan PaveSAM. Selain itu, model ini menunjukkan keupayaan generalisasi yang baik pada set data DeepCrack (skor Dice 74.7% dan IoU 62.43%), yang tidak pernah dilihat semasa latihan. PromptLessSAM menyediakan kaedah baharu yang cekap dan berkesan, serta menunjukkan cara menyesuaian model asas berskala besar bagi tugas khusus. Kajian ini

https://doi.org/10.31436/iiumej.v27i2.4124

menawarkan penyelesaian kukuh dengan prestasi tinggi namun dilatih menggunakan bilangan parameter paling rendah.

KEYWORDS: *Crack Semantic Segmentation, Segment Anything Model (SAM), Deep Learning, Structural Health Monitoring.*

1. INTRODUCTION

The regular inspection and maintenance of infrastructure, such as roads, bridges, and buildings, are essential for public safety and economic stability. This process is often called Structural Health Monitoring (SHM) [1]. Pavement distress, especially cracking, is a common and early sign of structural damage. Detecting these cracks early allows for timely repair, which can prevent bigger problems and save a lot of money [2]. For many years, this inspection was done manually by trained inspectors. This manual method is slow, costs a lot, can be dangerous for the inspectors, and the results can differ depending on the person [3].

To solve these problems, researchers have developed automatic methods using computer vision and deep learning [2]. Models like U-Net [4] and its variations have shown good results for semantic segmentation. This means assigning a label to every pixel in an image. More advanced models have also been used [5]. However, these deep learning models have two main challenges. First, they often need a very large number of parameters, making them heavy and slow. For example, some models use over 100 million parameters [6]. This makes them hard to use on small devices or for real-time inspection. Second, they need a large, high-quality dataset with pixel-level labels, which is very difficult and expensive to create [7].

Recently, a new type of model, called a "foundation model," has emerged. The Segment Anything Model (SAM) from Meta AI is a powerful example [8]. SAM was trained on a huge dataset of 11 million images and can segment almost any object in any image. The main problem is that SAM is a generalist model. It was not trained for specific tasks, such as pavement crack detection. As shown in other studies, SAM does not perform well on pavement cracks, which are very thin and have complex patterns, unlike the common objects in its training data [9]. Furthermore, SAM needs a "prompt" from the user, such as a point or a box, to determine what to segment. This makes it unsuitable for a fully automatic crack detection system.

Lightweight crack segmentation is important for real-world deployment because many inspection scenarios rely on edge platforms such as mobile robots, UAVs, or handheld devices, where memory, power, and latency are limited. Recent studies explicitly target this constraint; for example, CrackESS [10] validates crack segmentation in an edge setting and demonstrates deployment on a climbing robot platform using embedded hardware acceleration. Although foundation models such as SAM provide strong general representations, their large parameter size and compute cost make direct fine-tuning and deployment difficult in these constrained environments. Therefore, parameter-efficient adaptation is necessary because it can transfer foundation-model knowledge to the crack domain while updating only a small subset of parameters, thereby reducing training cost and making practical deployment more realistic.

This paper presents a method to solve these problems, called PromptLessSAM. We change the SAM foundation model from a generalist into a domain expert for crack segmentation. We do this by adapting it in a lightweight way. Our main idea is to keep most of the powerful, pre-trained image encoders of SAM frozen. This encoder already knows how to see rich features in an image. We then remove the original prompt encoder and mask decoder. We add our own new, very small decoder. These lightweight decoder and encoder neck layers are the only parts

https://doi.org/10.31436/iiumej.v27i2.4124

of the model that we train. This method is a form of Parameter-Efficient Fine-Tuning (PEFT) [11]. This has two benefits. First, it makes the model "prompt-less" because the decoder is trained to only find cracks. Second, it is efficient training. We only need to train about 1.1 million parameters. We show that our PromptLessSAM model outperforms many larger, more complex models on the Crack500 dataset. Our work provides an efficient way to adapt large foundation models for specific, important tasks.

The main contributions of this work are as follows:

- PromptLessSAM is proposed as a fully automated (prompt-less) crack segmentation approach by adapting the SAM image encoder with a lightweight crack-specific decoder.
- A parameter-efficient training strategy is applied by freezing the main encoder blocks and training only the encoder neck and decoder, which keeps the number of trainable parameters small.
- Competitive results are achieved on Crack500, and generalization to an unseen dataset (DeepCrack) is demonstrated without requiring any prompt input during inference.

The rest of this paper is organized as follows. Section 2 discusses the related work. Section 3 details our proposed methodology. Section 4 presents the experiments and results. Section 5 provides a comprehensive discussion of the findings and their significance. Finally, Section 6 concludes the paper and outlines directions for future research.

2. RELATED WORKS

The field of automated crack segmentation has seen much research, moving from traditional image processing to advanced deep learning [2]. Early methods used techniques like edge detection, thresholding, and morphological operations [12]. These methods are simple but are not robust. They fail in poor lighting, with shadows, or with different pavement textures. These traditional methods are often not reliable for complex road conditions [12]. In this review, representative works published are considered, covering CNN-based, Transformer-based, SAM-based, state-space (Mamba), and weakly/unsupervised approaches, and Table 1 summarizes their datasets, advantages, and limitations.

To overcome these issues, researchers turned to deep learning and convolutional neural networks (CNNs) [12]. The U-Net [4] architecture became very popular for this task. It uses an encoder-decoder structure with skip connections, which helps to get precise segmentation. Many other models are based on the U-Net. Other works, like DeepCrack [7], use a hierarchical feature learning approach. While these CNN-based methods perform well, they are often trained from scratch and require a large amount of labeled data to learn crack features.

More recently, Vision Transformers (ViT) [13] have shown great success in computer vision. Transformers use self-attention to capture long-range dependencies, which can be useful for segmenting long, thin cracks [14]. Models like TransUNet [6] combine the benefits of U-Net and Transformers. They use a Transformer as the encoder to learn global features and a CNN-based decoder to produce high-resolution segmentation [14]. However, these Transformer-based models are often very large, with many trainable parameters, like TransUNet, which has 101M parameters [6]. This makes them computationally expensive to train and deploy.

The release of the SAM [8] created a new direction for segmentation. SAM is a foundation model for segmentation. It can segment anything based on prompts. However, its zero-shot performance on specialized tasks like crack segmentation is poor [9]. Some recent research has

https://doi.org/10.31436/iiumej.v27i2.4124

tried to adapt SAM for specific domains. For example, PaveSAM [9] tried to use SAM for pavement distress but found its baseline performance was very low. CrackESS [10] proposed a system using YOLOv8 for self-prompting and a fine-tuned SAM. Other methods aim to make SAM smaller, such as FasterSAM [15]. Our work is different because we create a truly prompt-less model by replacing the decoder with a new, lightweight module, making it efficient and fully automatic.

Table 1. Related works summary (crack segmentation and SAM adaptation).

Model / Approach	Dataset Used	Key Finding	Advantage	Limitations / Evaluation metrics and accuracy
U-Net (CNN encoder–decoder) [4,18]	Crack500	Strong baseline with skip connections	Simple and stable	Full supervision; Eval: Precision (P), Recall (R), F1/Dice, IoU
DeepCrack (hierarchical CNN) [7]	Crack500, DeepCrack	Multi-level features improve crack detail	Good thin-crack sensitivity	Needs pixel labels; Eval: P, R, F1, IoU
TransUNet (Transformer + CNN) [6,19]	Crack500	Global context + decoder refinement	Better long-crack continuity	High compute; Eval: F1/Dice, IoU
SwinCrack (Swin Transformer) [18,20]	Crack500	Self-attention helps long-range crack modeling	High precision behavior	Heavier than CNN; Eval: P, R, F1, mIoU
VM-UNet32 (Mamba/SSM) [18,21]	Crack500	Mamba encoder improves context capture	Competitive performance	Complex design; Eval: P, R, F1, mIoU
DTrC-Net (hybrid) [19]	Crack500	Hybrid design for crack segmentation	Balanced accuracy/complexity	Higher params; Eval: mDS (Dice), mIoU
UP-CrackNet (unsupervised) [16]	Crack500, DeepCrack	Unsupervised via adversarial restoration	No pixel labels required	Sensitive to textures; Eval: Accuracy, P, R, F1, IoU
WP-CrackNet (weakly supervised) [17]	Crack500, DeepCrack	Weak supervision via adversarial learning	Reduced annotation cost	CAM incompleteness risk; Eval: Accuracy, P, R, F1, IoU
PaveSAM (SAM + box prompts) [9]	Crack500, Custom pavement distress dataset	SAM adaptation using box prompts for distress segmentation	Improves over zero-shot SAM	Requires prompts; Eval: P, R, F1, IoU
CrackESS (YOLO self-prompt + SAM) [10]	Crack500	Self-prompting (YOLO) + SAM-based segmentation for edge systems	Higher automation	Still prompt-based; Eval: P, R, F1, IoU, FPS

In addition to fully supervised CNN and Transformer models, recent studies investigate methods to reduce dependence on expensive pixel-level labels. UP-CrackNet [16] introduces an unsupervised pixel-wise framework that uses adversarial image restoration to learn crack features without direct crack supervision. WP-CrackNet [17] extends this line of work by using a collaborative adversarial learning framework for end-to-end weakly supervised road crack detection, showing that weaker annotations can still yield usable segmentation outputs. Furthermore, lightweight designs based on state-space models (SSMs) and Mamba have been explored to capture long-range crack patterns at reduced computational cost compared with heavy Transformers [18,19]. Table 1 summarizes representative recent studies, including

https://doi.org/10.31436/iiumej.v27i2.4124

datasets, key findings, advantages, and limitations, and evaluates them using the evaluation metrics to clarify the research gap addressed by our prompt-less SAM adaptation.

3. METHODOLOGY

In this section, we first revisit SAM and then introduce the task addressed in this paper: SAM for Pavement Distress Segmentation. Next, we delve into how to achieve the pipeline of the proposed framework. Finally, we outline the training and evaluation processes employed in our approach.

3.1. Preliminary

3.1.1. Revisit of Segment Anything Model

The SAM [9] is a foundation model designed for promptable segmentation. It was trained on the large SA-1B dataset. This training allows SAM to have strong zero-shot generalization capabilities. As shown in Figure 1, SAM's architecture has three main parts. First is a powerful image encoder, which is a ViT [8]. This encoder processes the input image and creates a rich feature embedding. Second is a prompt encoder. This part takes user inputs, such as points, boxes, or text, and converts them into an embedding. Third is a lightweight mask decoder. This decoder takes the image embedding and the prompt embedding and produces the final segmentation mask. This design is flexible, but it always requires a prompt to tell the model what to segment.

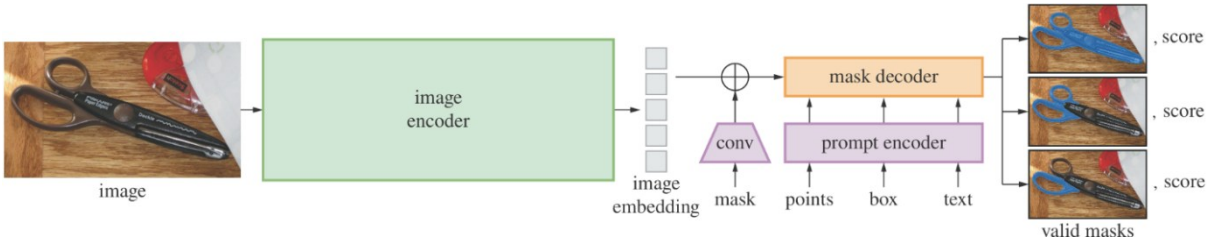


Figure 1. The architecture of the original SAM, showing the Image Encoder, Prompt Encoder, and Mask Decoder.

3.1.2. SAM on Pavement Distress Segmentation

While SAM is effective for general segmentation, it struggles with specialized tasks such as pavement crack segmentation. This is because pavement distresses are very different from the "natural" objects in SAM's training data [8]. Cracks are often very thin, have low contrast with the road surface, and form complex, irregular networks [7]. When SAM is applied to crack images without any task-specific training, its performance is very poor.

Research by Owor et al. on PaveSAM [9] evaluated SAM's zero-shot performance on crack datasets. Their results showed that SAM alone is not suitable for this task. For example, on the Crack500 dataset, their experiments showed that SAM achieved a Dice score of only 0.137. This is much lower than even older models such as U-Net [4]. This poor performance shows that SAM, in its original form, cannot be used as a reliable tool for automatic pavement inspection. It does not understand the specific features of cracks. Therefore, adaptation is necessary to make it useful for this domain. SAM does not produce good results on pavement distress, as shown in Figure 2.

https://doi.org/10.31436/iiumej.v27i2.4124

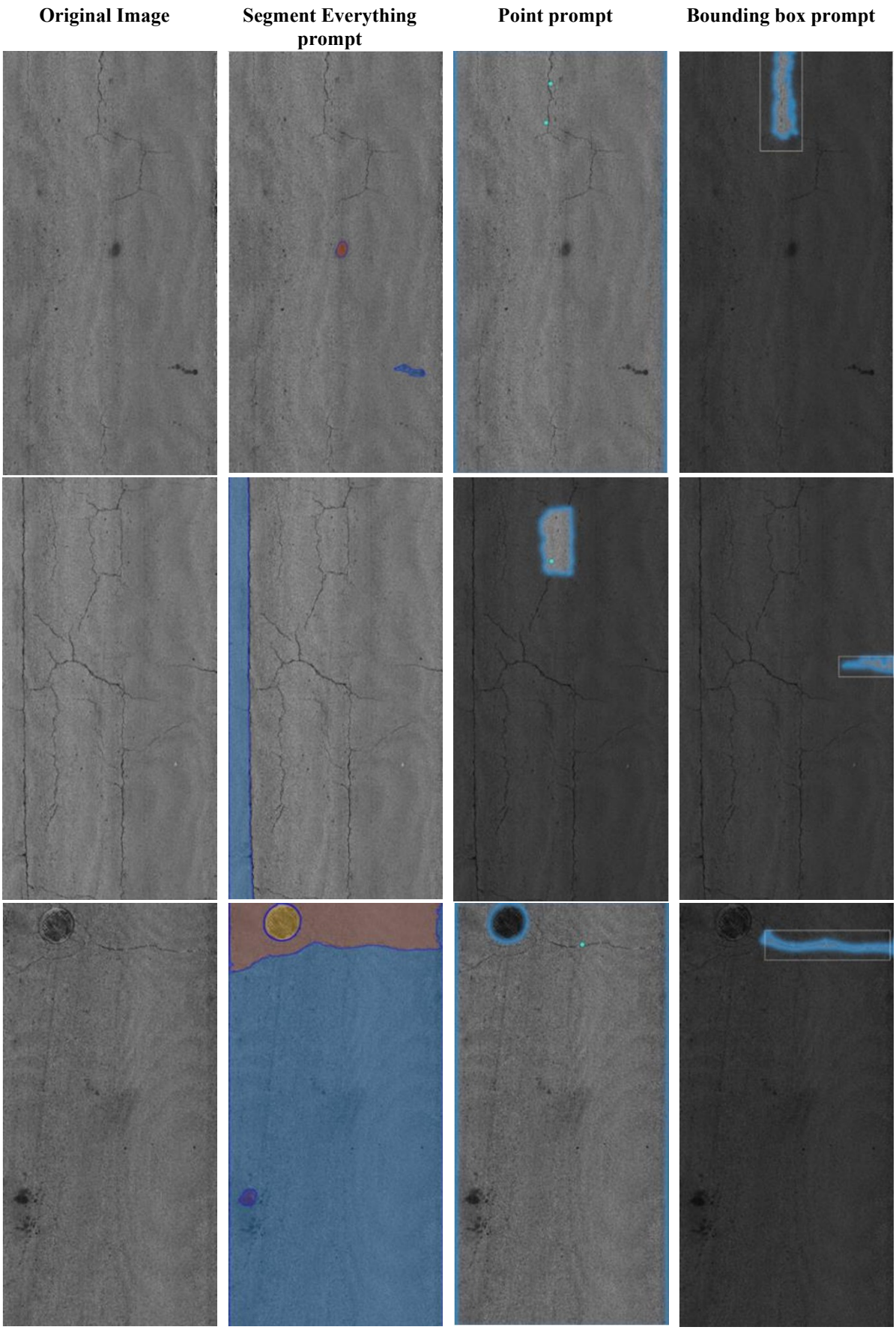


Figure 2. Examples of SAM's poor zero-shot performance on pavement crack images [9].

https://doi.org/10.31436/iiumej.v27i2.4124

3.2. Proposed Method (PromptLessSAM)

To address poor performance and the need for prompts, we propose PromptLessSAM. Our goal is to transform SAM from a general-purpose, promptable model into an automatic, specialized "expert" for crack segmentation. We do this by changing its architecture. The key idea is to leverage the powerful pre-trained SAM image encoder while replacing the other parts. The architecture of our proposed model is shown in Figure 3.

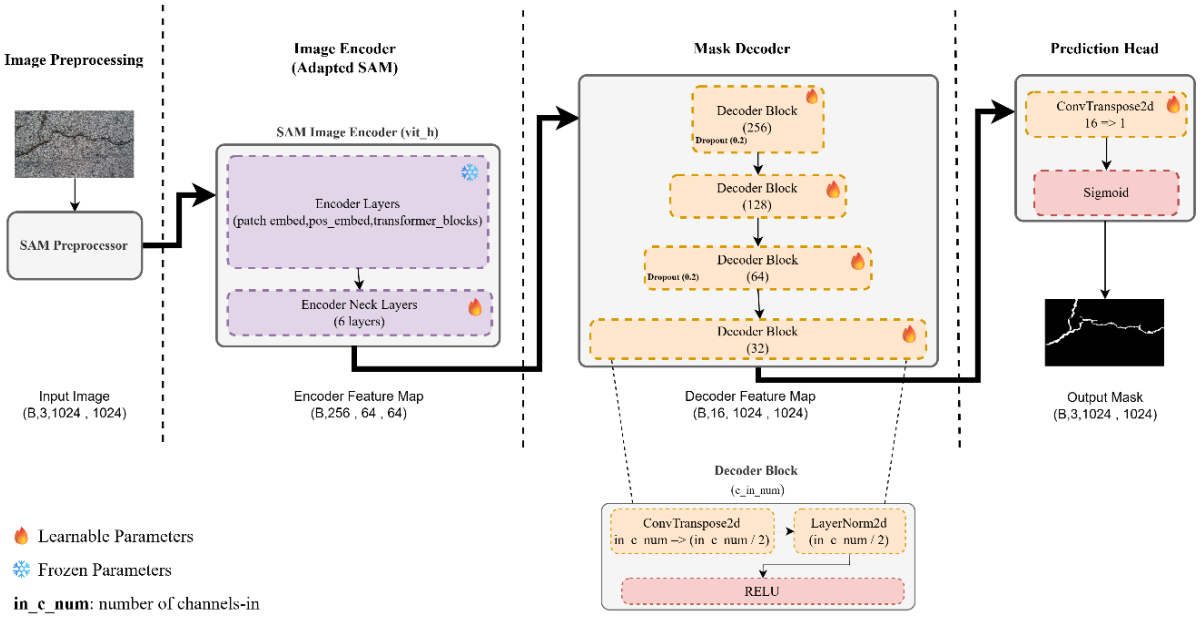


Figure 3. The proposed architecture of PromptLessSAM.

3.2.1. Model Framework Overview

Our framework has two main components: a partially frozen image encoder and a lightweight, trainable decoder. We take the pre-trained image encoder from SAM. This encoder comprises a main ViT-H backbone and a "Neck" module. We freeze the ViT-H backbone weights. This is important because this part has billions of parameters and has already learned powerful, general-purpose features. By freezing it, we save a huge amount of computation and prevent overfitting.

We then allow the "Neck" of the SAM encoder to be trainable. This part of the encoder helps to process the features from the ViT backbone. By fine-tuning it, we adapt these general features to the specific task of crack detection. We completely remove SAM's original prompt encoder and mask decoder. In their place, we attach a new, lightweight decoder. This decoder is also trainable and learns to interpret the adapted features from the Neck to find *only* cracks. Because the Neck and decoder are trained only on crack images, the model automatically outputs a crack mask for any input image. This makes our model fully automatic.

3.2.2. Image Preprocessing

To use the powerful SAM image encoder, it is necessary to prepare the input images correctly. The encoder is a ViT and was pre-trained on images of a specific size and format. If we give the encoder different images, it will not produce good features. Therefore, our model uses the standard preprocessing function from the original SAM [8]. Preprocessing resizes every input image from the dataset to a fixed resolution of 1024x1024 pixels. After resizing, the function also normalizes the pixel values using the mean and standard deviation used by

https://doi.org/10.31436/iiumej.v27i2.4124

the SAM model during training. This step is very important to make sure all images are in the correct format for the frozen ViT backbone.

3.2.3. PromptLessSAM Architecture

Our proposed architecture comprises a modified SAM encoder and a new lightweight decoder, described in detail in this section.

a. Image Encoder

The image encoder is the backbone of our model. We use the ViT-H encoder from the original SAM. This encoder, as shown in Figure 3, can be divided into two parts: the main Image Encoder (ViT-H) backbone and a Neck. The ViT-H backbone processes the input image using a series of Transformer blocks to create a rich feature embedding [13]. In our method, this large backbone is completely frozen.. Its weights are not updated during training.

The output of the frozen ViT-H backbone is then passed into the Neck. The Neck job refines features from the backbone. Unlike the backbone, we make this Neck trainable. This allows the model to learn to adapt general features to the specific, fine-grained details of cracks. The output of this trainable Neck is then passed to our new decoder.

b. Decoder Block

The Decoder Block is the fundamental building block of our lightweight decoder. Its main purpose is to take the encoder's feature maps and progressively upsample them to reconstruct the full-resolution mask. As shown in Figure 3, each Decoder Block is a sequence of layers. The block starts with a Transposed Convolution (ConvTranspose2d) layer. This layer performs upsampling, which doubles the spatial dimensions of the feature map (e.g., from 64×64 to 128×128) while reducing the number of feature channels. The transposed convolutions use a 2×2 kernel with a stride of 2 and no padding.

After the upsampling, we apply a LayerNorm2d layer. This is the same 2D layer normalization used in the original SAM architecture, which normalizes the features to stabilize training. This is followed by a Rectified Linear Unit (ReLU) activation function, which introduces non-linearity to help the model learn complex crack patterns. A dropout layer is also added to some of the blocks. This is a regularization technique that helps prevent overfitting during training.

c. Lightweight Decoder

The core of our contribution is the lightweight decoder, which is built by stacking the Decoder Blocks described in the previous section. This decoder is designed to be small and efficient. It has a U-Net-like structure that takes the low-resolution feature map from the encoder's trainable Neck and gradually upsamples it. The SAM encoder output is 256×64×64. Our decoder uses four Decoder Blocks in sequence.

These blocks double the resolution and halve the channels at each step (256×64×64 → 128×128×128 → ... → 16×512×512).

d. Prediction Head

After the last Decoder Block, the feature map has a resolution of 1024x1024 with 16 feature channels. The final part of our network is the Prediction Head. This part converts these features into the final segmentation mask. It consists of a single 2D transposed convolution layer, which acts as a 1x1 convolution. This layer has a kernel size of 1 and a stride of 1.

https://doi.org/10.31436/iiumej.v27i2.4124

The purpose of this layer is to reduce the number of feature channels from 16 to 1. This operation creates a single-channel feature map. This map is then passed through a Sigmoid activation function. The Sigmoid function is important because it maps all values to the [0,1] interval. The result is the final probability map, in which each pixel value represents the model's confidence that the pixel belongs to a crack.

3.3. Dataset

We used the public Crack500 dataset [9] for training and testing our model. This dataset is a standard benchmark for pavement crack segmentation, as used in many studies [9], [10]. It contains 500 images of size around 2000 1500 pixels of pavement cracks. The images are high-resolution and cover a range of crack types and pavement conditions. The dataset is divided into 250 training images, 50 validation images, and 200 test images. It includes high-quality pixel-level ground-truth masks. We also use the DeepCrack dataset [7] to test our model's generalization ability. The dataset consists of 537 pixel-level annotated images of cracks on concrete and asphalt surfaces across various scenes and scales, with a resolution of 544 × 384 pixels [17]. In a common split, 300 images are used for training, and 237 images are used for testing [16]. In our experiments, we do not train or tune the model on DeepCrack and use it only as an unseen test set to evaluate cross-dataset generalization.

3.4. Training Strategy

3.4.1. Implementation

Our model was implemented using the PyTorch [22] library and trained on a cloud platform (RunPod) using NVIDIA GPU (RTX 5090). The model uses the Adam [23] optimizer with beta values of (0.9, 0.999).

Our training process was done in steps. We first trained the model using binary cross-entropy (BCE) with a learning rate of 0.01. We found that the best model checkpoint on the testing dataset was achieved at epoch 12. We then reloaded this best checkpoint (epoch 12) and continued training for another 25 epochs. For this second phase of training, we used Dice-BCE loss and the same learning rate (0.01) to further fine-tune the model's performance.

3.4.2. Loss Function

Choosing the right loss function is very important for segmentation tasks, especially with imbalanced data. Crack segmentation is highly imbalanced because most of the image is "not crack" (background) and only a few pixels are "crack".

Binary Cross-Entropy (BCE) [24] Loss is a popular loss function; it performs well for pixel-level classification. BCE loss is used to minimize pixel-level errors by penalizing the difference between predicted probabilities and the actual class labels, encouraging accurate predictions. BCE loss is defined as:

$$Loss_{BCE}(y, \hat{y}) = -(y\, log(\hat{y}) + (1 - y)\, log(1 - \hat{y})) \quad (1)$$

where $\hat{y}$ is the predicted value by the prediction model, and y is the actual value.

Dice Loss [25] is very good for imbalanced data because it measures the overlap between the predicted mask and the true mask. The Dice coefficient measures this overlap. Dice loss is defined as follows:

$$Loss_{Dice} = 1 - \frac{2\Sigma_{i=1}^{n} p_i y_i}{\Sigma_{i=1}^{n} p_i^2 + \Sigma_{i=1}^{n} y_i^2} \quad (2)$$

where y_i is the actual pixel value, and p_i is the predicted pixel value.

https://doi.org/10.31436/iiumej.v27i2.4124

Our final loss function is the sum of these two losses:

$$Loss_{final} = Loss_{BCE} + Loss_{Dice} \quad (3)$$

This combined loss helps the model produce accurate pixel classifications (from BCE) while also ensuring high overlap with the true cracks (from Dice Loss).

3.4.3. Evaluation Methods

To evaluate the performance of the PromptLessSAM model, we use several standard semantic segmentation metrics. These metrics are calculated using the True Positives (TP), False Positives (FP), and False Negatives (FN) from the test set.

- Precision [26]: This measures how many of the pixels we predicted as "crack" were actually cracks.

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

- Recall [27]: This measures how many of the "crack" pixels in the image we were able to find.

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

- Dice Score (F1-Score) [7]: This is the harmonic mean of Precision and Recall. It is a very common metric for segmentation.

$$F1\ score = \frac{2*Precision*Recall}{Precision+Recall} \quad (6)$$

- Intersection over Union (IoU) [26]: This is another very common metric that measures the overlap between the predicted and true masks.

$$IoU = \frac{TP}{TP+FP+FN} \quad (7)$$

We use these four metrics to compare our model to other methods.

4. EXPERIMENTS AND RESULTS

4.1. Implementation Details

As mentioned in Section 3.4.1, the model was built in PyTorch [22] and trained using an Adam optimizer [23]. The Learning Rate (LR) was set to 0.01 because this value is commonly used as an initial LR in crack segmentation training, especially when the training schedule is short or when only part of the model is optimized. For example, the importance-enhanced Mamba-based crack segmentation network reports using the Adam optimizer with an initial learning rate of 0.01 during training. [18] In addition, UP-CrackNet [16] reports an initial learning rate of 0.01 for supervised crack segmentation settings (with poly adjustment), indicating that this LR scale is practical for crack segmentation optimization. In our case, only the encoder neck and lightweight decoder are trainable, so a moderate LR helps the model converge efficiently. A much smaller LR reduces the update magnitude and can slow convergence.

The training was performed on a RunPod [28] cloud computing platform, which provided the necessary GPU resources. The model uses a frozen SAM ViT image encoder and a custom lightweight decoder. The total number of trainable parameters in our model is only 1,093,345 (~1.1 million).

https://doi.org/10.31436/iiumej.v27i2.4124

4.2. Qualitative Results and Visual Comparison

We compare the performance of our PromptLessSAM model with several other well-known methods for crack segmentation. We chose these methods for comparison for specific reasons. We compare against the U-Net [4] as a baseline CNN architecture. We include Deepcrack [7] and TransUNet [6] because they are popular, high-performance models, but they have a very large number of trainable parameters (14.72M and 101.19M, respectively). We also compare against SAM-based methods like PaveSAM [9] and CrackESS [10] because they also use the SAM foundation model, but in a different way.

Table 2 shows the quantitative results on the Crack500 dataset. PromptLessSAM achieves a high Dice score of 72.7% and an IoU of 57.11%, which outperforms U-Net [4], Deepcrack [7] and TransUNet [6,18]. It also surpasses the SAM-based methods (SAM [8], PaveSAM [9], and CrackESS [10]), demonstrating that prompt-less adaptation can be effective. In addition, we include more recent crack segmentation networks with lower reported results on Crack500, such as UP-CrackNet [16] and WP-CrackNet [17], as well as SwinCrack [18,21], VM-UNet32 [18,21], and DTrC-Net [19], to provide a broader comparison. Overall, PromptLessSAM remains the best-performing model in terms of recall and Dice score.

The other important result is the comparison of performance and model size. PromptLessSAM achieves the best segmentation performance while having the fewest trainable parameters (1.1M). This has almost 100 times fewer parameters than TransUNet [6,18] (101.19M) and 13 times fewer than Deepcrack [7] (14.72M). This shows that our method is extremely training-efficient.

Table 2. Evaluation metrics of competing methods on the Crack500 dataset.

Methods	Precision [%]	Recall [%]	F1 [%]	IoU [%]	Trainable Params [M]
U-Net [4]	63.90	68.40	65.10	–	13.34
Deepcrack [7]	62.80	66.10	67.60	–	14.72
TransUNet [6,18]	60.50	69.60	64.70	–	101.19
SwinCrack [18,20]	**73.38**	68.26	70.73	**72.77**	116.23
VM-UNet32 [18,21]	70.27	72.35	71.69	–	27
DTrC-Net [19]	–	–	67.50	53.30	41.8
UP-CrackNet [16]	65.38	58.61	61.81	44.73	–
WP-CrackNet [17]	71.81	55.58	62.66	45.63	–
CrackESS (4-point) [10]	51.70	70.50	56.10	40.70	–
SAM [9]	10.50	67.20	14.80	9.50	–
PaveSAM (Dice + BCE loss) [9]	69.60	72.30	69.10	53.80	3.87
PromptLessSAM (Ours)	70.25	**75.32**	**72.70**	57.11	**1.09**

A graph was used to demonstrate the quantitative comparison of models' performance based on the evaluation metrics mentioned above. Figure 4 shows that our model achieved the highest recall (75.32%) and the highest Dice score (72.70%) among state-of-the-art models. At the same time, SwinCrack [18,20] achieved the highest precision and IoU among the reported results, indicating strong overlap performance with a large model. However, when considering the trainable parameters in Table 2, PromptLessSAM provides a better trade-off between segmentation accuracy and training cost. This confirms that lightweight decoder adaptation can still achieve competitive crack segmentation without relying on prompts.

https://doi.org/10.31436/iiumej.v27i2.4124

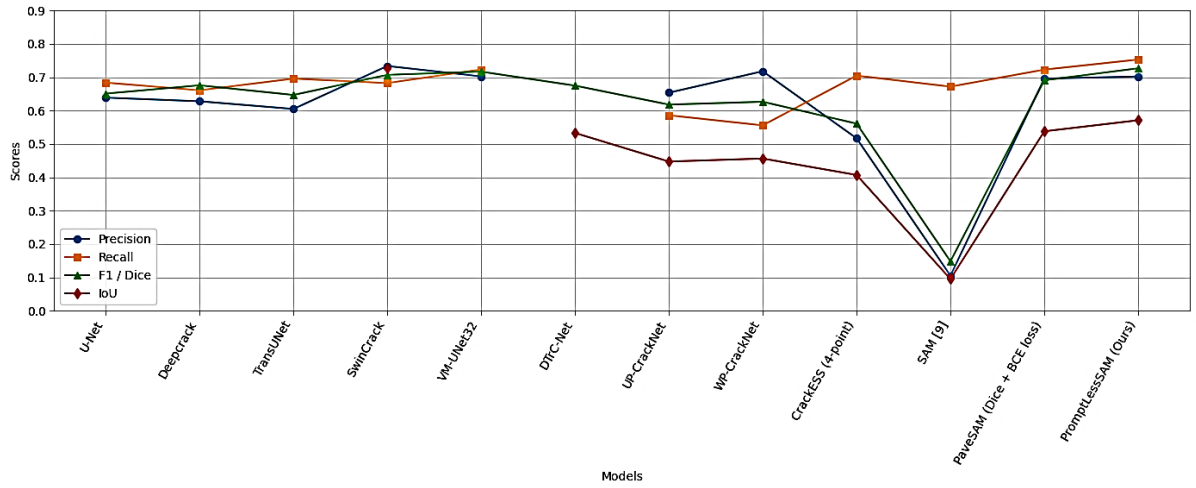
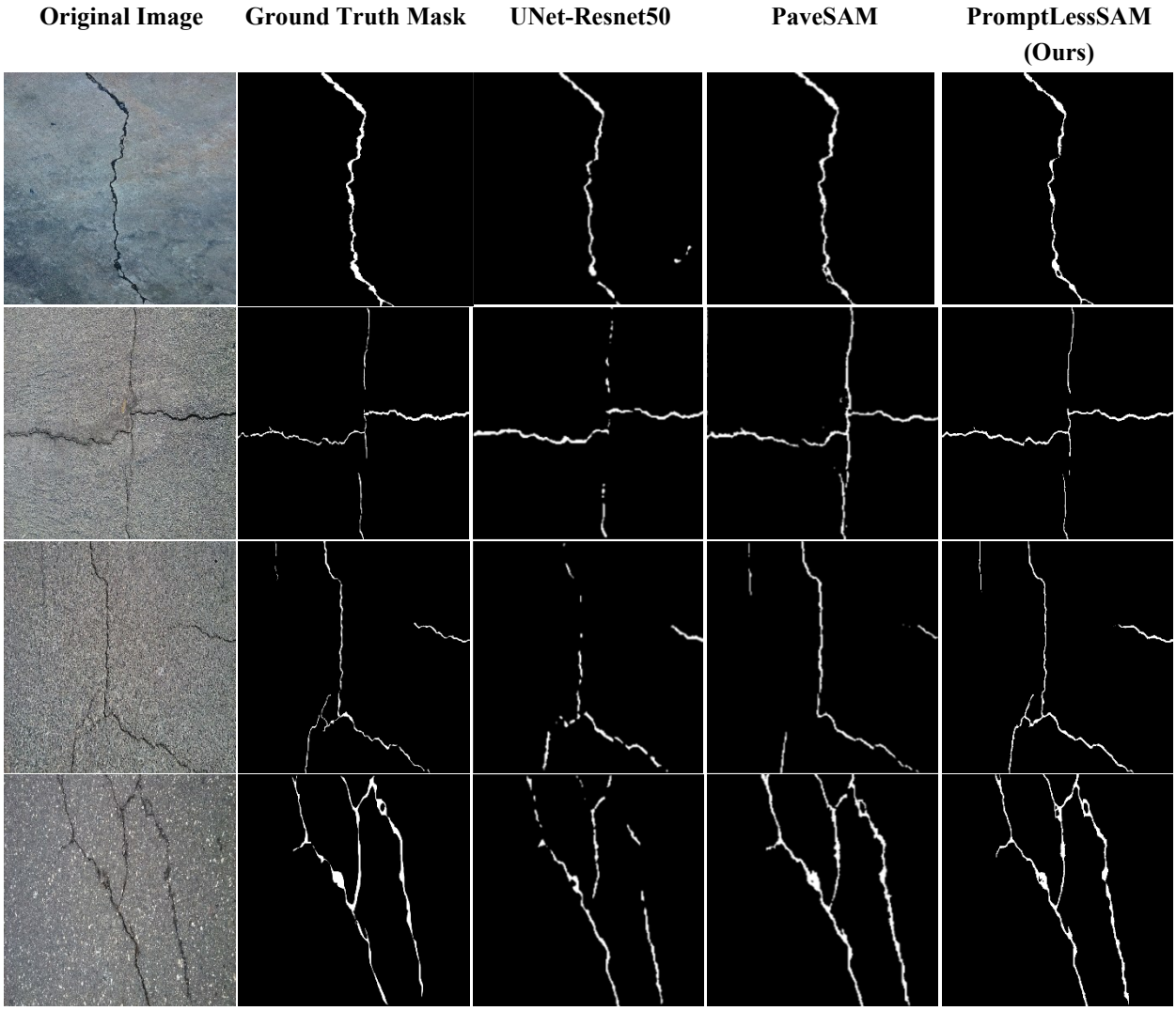


Figure 4. Quantitative comparison of performance metrics across different segmentation models.

Visual examples of our model's output on the Crack500 dataset (test data) are illustrated in Figure 5.



https://doi.org/10.31436/iiumej.v27i2.4124

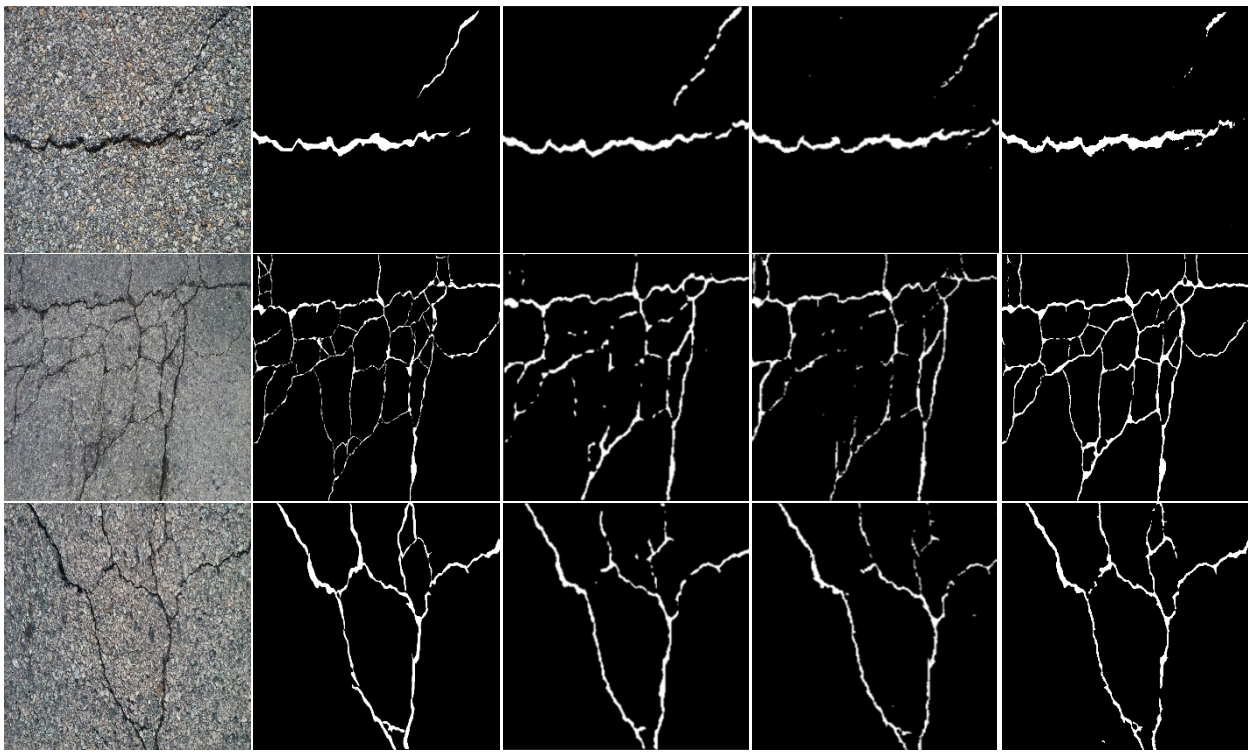


Figure 5. Visual comparison of segmentation results on the Crack500 test set.

The visual results demonstrate that our model generalizes well to different crack shapes and backgrounds, such as asphalt and concrete, in the dataset.

4.3. Generalization on DeepCrack Dataset

To check if our model simply memorized the Crack500 dataset, we tested its generalization ability. We took our trained PromptLessSAM model (which was trained only trained on Crack500) and tested it directly on the DeepCrack dataset [7]. The model had never seen any images from this dataset before.

The quantitative evaluation on this unseen dataset demonstrates the robust performance of our method. Specifically, PromptLessSAM achieved a precision of 79.32% and a recall of 78.25%. Furthermore, the model achieved a strong F1 Score of 74.70% and an IoU of 62.43%. This performance is slightly better than its F1 score on the Crack500 dataset (73.3%). This is an important result because it shows that the model is not overfitting to the training data. This result demonstrates that our fine-tuning strategy has learned a robust, general representation of what a “crack” is and can be applied to new, unseen datasets. This proves that our method is a reliable and general-purpose solution for crack segmentation. Visual examples of our model’s output on the DeepCrack dataset are shown in Figure 6.

https://doi.org/10.31436/iiumej.v27i2.4124

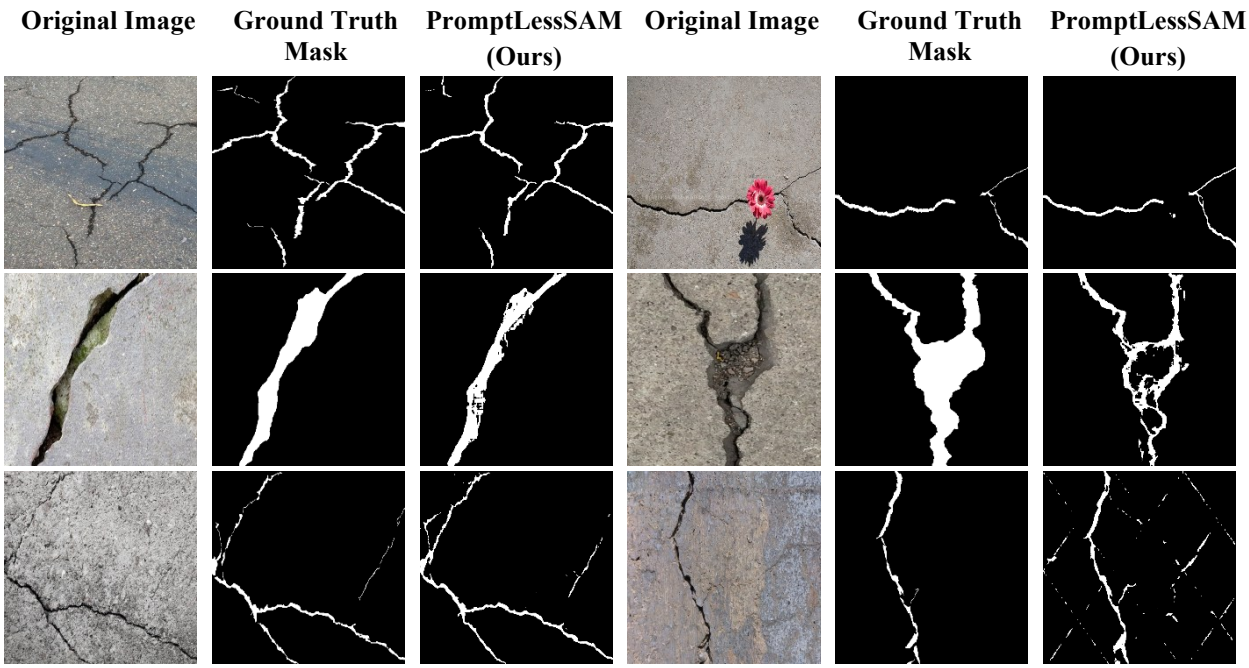


Figure 6. Visual examples of our model's generalization performance on the DeepCrack dataset (never seen).

5. DISCUSSION

PromptLessSAM demonstrates that a large pre-trained model can be effectively repurposed for a specialized task with minimal training. By removing the dependency on prompts, we simplify deployment; an inspector using our model does not need to provide any hint; the model directly outputs crack locations. This contrasts with PaveSAM, which requires drawing a box around each distress [9]. Our approach can thus be integrated into autonomous inspection systems, such as those mounted on drones or robots, to enable fully automated crack mapping.

The Dice score of 72.7% on Crack500 should be interpreted in the context of dataset difficulty. Crack500 includes challenging imaging conditions such as shadows, occlusions, varying lighting, and strong background interference, which makes accurate segmentation of thin cracks harder [17,18]. For this reason, many published methods on Crack500 commonly report Dice/F1 values in the 60% to low 70% range (for example, WP-CrackNet [17] and UP-CrackNet [16] are around 62%, while SwinCrack [20] and VM-UNet32 [18] are around 71%). However, on other crack datasets with different characteristics, several models can reach close to 80% F1, such as the reported results on Crack756 in recent Mamba-based studies [18]. Therefore, our Crack500 result indicates that the proposed prompt-less SAM adaptation handles a hard benchmark while keeping the adaptation lightweight.

In addition, our implementation was initially based on an open-source promptless task-specific fine-tuning pipeline [29] built on top of the SAM framework. This reference work provides a practical baseline for adapting SAM to a single target class without interactive prompting, aligning with our goal of fully automatic crack segmentation. However, our final model differs in several important aspects to better fit the crack segmentation problem, especially by changing the loss functions and tuning various training parameters to address the severe class imbalance between crack pixels and background. Therefore, while we acknowledge this codebase as a starting point, our modifications are essential to improve performance and to ensure the contribution of our proposed method to pavement crack datasets.

One limitation is that SAM's encoder operates at a fixed input size (1024×1024). For very large images, one must tile or resize, which could affect very small cracks. Another point is inference speed: while adequate models achieve much higher speed, there is a trade-off between segmentation quality and speed. For real-time requirements, one might opt for a smaller SAM or distill our model.

In terms of broader impact, our work reinforces the idea that foundation models can serve as powerful backbones for domain-specific tasks in civil engineering. The success of PromptLessSAM suggests that training task-specific models from scratch might become unnecessary when such pre-trained models are available. This can significantly reduce development time and the amount of required training data for new vision tasks in SHM and beyond.

6. CONCLUSION

In this paper, we propose PromptLessSAM, an efficient method for adapting the large SAM foundation model to the specific task of pavement crack segmentation. We addressed the main problem of existing methods: the poor domain-specific performance of generalist models like SAM. Our method freezes the powerful SAM ViT-H backbone, fine-tunes the encoder's Neck, and trains a new lightweight decoder. This adaptation makes the model "prompt-less" and fully automatic.

Our experiments on the Crack500 dataset show that our model is extremely efficient, with only 1,093,345 (~ 1.1 million) trainable parameters. With this lightweight design, our model achieves a Dice score of 72.7%. This performance is higher than that of many other popular models. Furthermore, our model showed good generalization, achieving a 74.7% Dice score on the unseen DeepCrack dataset. This work shows that this lightweight adaptation strategy is highly effective and efficient for turning foundation models into domain experts.

For future work, several clear directions are planned. First, we aim to improve the Dice similarity coefficient by using stronger training protocols and greater data diversity, including additional augmentations, longer training schedules, and patch-based training and inference to refine thin-crack boundaries and improve continuity. Second, deployment aspects should also be addressed by reducing encoder cost through model compression, including quantization. Finally, acceleration and optimized export for mobile and edge devices will be explored, following systems such as CrackESS.

REFERENCES

[1] Sundravel KV, Jagadeesan R, Dhanush A, Kumar AS (2025). A Review of Innovations, Challenges, and Future Directions in Structural Health Monitoring Using Smart Materials. Institute for Environmental Nanotechnology. doi:10.13074/jent.2025.03.2511163.

[2] Zhang X, Wang H, Hsieh YA, Yang Z, Yezzi A, Tsai YC (2025). Deep Learning for Crack Detection: A Review of Learning Paradigms, Generalizability, and Datasets. Retrieved from http://arxiv.org/abs/2508.10256

[3] Benmhahe B, Chentoufi JA (2021). Automated Pavement Distress Detection, Classification and Measurement: A Review. International Journal of Advanced Computer Science and Applications, 12(8):708–718. https://doi.org/10.14569/IJACSA.2021.0120882

[4] Ronneberger O, Fischer P, Brox T (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp 234–241. Retrieved from http://arxiv.org/abs/1505.04597

[5] Chen X, Liu C, Chen L, Zhu X, Zhang Y, Wang C (2024). A Pavement Crack Detection and Evaluation Framework for a UAV Inspection System Based on Deep Learning. Applied Sciences, 14(3). doi:10.3390/app14031157.

[6] Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, Lu L, Yuille AL, Zhou Y (2021). TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. Retrieved from https://arxiv.org/pdf/2102.04306

[7] Liu Y, Yao J, Lu X, Xie R, Li L (2019). DeepCrack: A deep hierarchical feature learning architecture for crack segmentation. Neurocomputing, 338:139–153. doi:10.1016/j.neucom.2019.01.036.

[8] Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo WY, Dollár P, Girshick R (2023). Segment Anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp 4015-4026. Retrieved from http://arxiv.org/abs/2304.02643

[9] Jakisa N, Student OP, Adu-Gyamfi Y, Aboah A, Amo-Boateng M (2024). PaveSAM-Segment Anything for Pavement Distress.

[10] Wang Y, He J, Yu S (2025). CrackESS: A Self-Prompting Crack Segmentation System for Edge Devices. Retrieved from http://arxiv.org/abs/2412.07205

[11] Zhang D, Feng T, Xue L, Wang Y, Dong Y, Tang J (2025). Parameter-Efficient Fine-Tuning for Foundation Models. Retrieved from http://arxiv.org/abs/2501.13787

[12] Huang S, Chen H, Yan L, Zou X, Li B, Bi Y (2025). A review of the progress in machine vision-based crack detection and identification technology for asphalt pavements. Digital Transportation and Safety, 4(1):65–79. doi:10.48130/dts-0025-0006.

[13] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In International Conference on Learning Representations (ICLR). Retrieved from http://arxiv.org/abs/2010.11929

[14] Li X, Ding H, Yuan H, Zhang W, Pang J, Cheng G, Chen K, Liu Z, Loy CC (2024). Transformer-Based Visual Segmentation: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(12):10138-10159. Retrieved from http://arxiv.org/abs/2304.09854

[15] Zhao X, Ding W, An Y, Du Y, Yu T, Li M, Tang M, Wang J (2023). Fast Segment Anything. Retrieved from http://arxiv.org/abs/2306.12156

[16] Ma N, Fan R, Xie L. (2024). UP-CrackNet: Unsupervised Pixel-Wise Road Crack Detection via Adversarial Image Restoration. IEEE Transactions on Intelligent Transportation Systems, 25(10):13926–13936. doi: 10.1109/TITS.2024.3398037

[17] Ma N, Song Z, Hu Q, Tang X, Zhang C, Fan R, Xie L. (2026). WP-CrackNet: A collaborative adversarial learning framework for end-to-end weakly-supervised road crack detection. Neurocomputing, 659:131845. doi: 10.1016/j.neucom.2025.131845

[18] Wang Y, Jin J, Chen X, Wu Z, Zhang L. (2025). A lightweight crack segmentation network based on the importance-enhanced Mamba model. Scientific Reports, 15(1):25504. doi: 10.1038/s41598-025-25504-4

[19] Tang W, Wu Z, Wang W, Pan Y, Gan W. (2025). VM-UNet++ research on crack image segmentation based on improved VM-UNet. Scientific Reports, 15(1):92994. doi: 10.1038/s41598-025-92994-7

[20] Wang C, Liu H, An X, Gong Z, Deng F. (2024). SwinCrack: Pavement crack detection using convolutional swin-transformer network. Digital Signal Processing, 145:104297. doi: 10.1016/j.dsp.2023.104297

[21] Ruan J, Li J, Xiang S. (2025). VM-UNet: Vision Mamba UNet for Medical Image Segmentation. ACM Transactions on Multimedia Computing, Communications, and Applications, 20(23):1-15. doi: 10.1145/3767748

[22] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Köpf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. Retrieved from http://arxiv.org/abs/1912.01703

[23] Kingma DP, Ba J (2017). Adam: A Method for Stochastic Optimization. Retrieved from http://arxiv.org/abs/1412.6980

[24] Jadon S (2020). A survey of loss functions for semantic segmentation. In IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). doi:10.1109/CIBCB48159.2020.9277638.

[25] Rajput V (2021). Robustness of different loss functions and their impact on network's learning capability. Retrieved from https://arxiv.org/abs/2110.08322

[26] Hussan PH, Ali IH (2025). ECGANCOVID: Efficient Conditional GAN Architecture for Covid-19 Disease Segmentation. Baghdad Science Journal, 22(2):706–729. doi:10.21123/bsj.2024.9335.

[27] Al Akabi H, Al-Assadi T (2025). CLASSIFICATION OF FAULT SIGNALS BASED ON DCT AND DEEP LEARNING. Kufa Journal of Engineering, 16(4):217–234. doi:10.30572/2018/KJE/160413.

[28] Runpod. Runpod | The cloud built for AI. Retrieved from https://www.runpod.io/

[29] Yatnalkar Y (2025). SAM-Promptless-Task-Specific-Finetuning. Retrieved from https://github.com/yogendra-yatnalkar/SAM-Promptless-Task-Specific-Finetuning

https://doi.org/10.31436/iiumej.v27i2.4240

Stability-Aware Evaluation of a CNN–LSTM–DQN Intrusion Detection System for Zero-Day and Drifted Network Traffic

RUSHENDRA, KALAMULLAH RAMLI*, PRIMA DEWI PURNAMASARI

Department of Electrical Engineering, Universitas Indonesia, Depok, Indonesia

**Corresponding author: kalamullah.ramli@ui.ac.id*

(Received: 26 February 2026; Accepted: 30 April 2026; Published online: 10 May 2026)

ABSTRACT: Intrusion Detection Systems (IDS) deployed in real-world environments must operate under severe class imbalance, evolving attack strategies, and non-stationary traffic distributions. Conventional supervised and deep learning–based IDS rely on fixed decision functions, limiting their adaptability to zero-day attacks and concept drift. This paper proposes a hybrid CNN–LSTM–DQN framework combined with a stability-aware evaluation methodology. The CNN–LSTM backbone extracts spatio-temporal representations, while a Deep Q-Network (DQN) learns adaptive detection policies using an ARMF-aware reward formulation. The framework is evaluated on eleven experimental stages (E1-11), including supervised baselines, reinforcement learning optimization, zero-day generalization (LOAO), and drift scenarios. Experimental results show that supervised models have high recall (up to 98.39%) but generate too many alerts (ARMF up to 44,845). The reinforcement learning model with prioritized experience replay (E7) achieves a more balanced performance with a recall of 91.40% and an ARMF of 1,031. The proposed PER-based approach significantly improves detection performance while maintaining low alert rates, achieving a recall of 42.47% compared to naive reinforcement learning (E5). Further evaluation in real drifting conditions showed robust recall (89-92%) with a tolerable number of alerts (ARMF ≈ 1,189). These results indicate that adaptive policy learning enables a more effective trade-off between detection performance and operational cost, while ARMF-based evaluation provides a practical complement to accuracy metrics for real-world IDS deployment.

ABSTRAK: Sistem Pengesanan Pencerobohan (IDS) dalam dunia nyata perlu beroperasi di bawah ketidakseimbangan kelas yang teruk, strategi serangan berevolusi, dan taburan trafik tidak pegun. IDS berasaskan pembelajaran penyeliaan dan pembelajaran mendalam konvensional bergantung pada fungsi keputusan tetap, mengehadkan kebolehsuaian terhadap serangan sifar hari dan hanyutan konsep. Kajian ini mencadangkan gabungan rangka kerja hibrid CNN–LSTM–DQN dan metodologi penilaian yang mementingkan kestabilan. CNN–LSTM mengekstrak representasi ruang-masa, manakala Rangkaian Q Mendalam (DQN) mempelajari dasar pengesanan adaptif menggunakan formulasi ganjaran ARMF. Rangka kerja ini dinilai dengan sebelas peringkat eksperimen (E1-11) termasuk garis dasar penyeliaan, pengoptimuman pembelajaran pengukuhan, generalisasi sifar hari (LOAO) dan senario hanyutan. Dapatan eksperimen menunjukkan bahawa model penyeliaan mempunyai kadar ingat semula (recall) yang tinggi (sehingga 98.39%) tetapi menjana terlalu banyak amaran (ARMF sehingga 44,845). Model pembelajaran pengukuhan dengan ulangan pengalaman berprioriti (E7) mencapai prestasi lebih seimbang dengan kadar ingat semula 91.40% dan ARMF sebanyak 1,031. Pendekatan berasaskan PER yang dicadangkan meningkatkan keupayaan pengesanan dengan ketara sambil mengekalkan kadar amaran yang rendah berbanding pembelajaran pengukuhan naif (E5) dengan kadar ingat semula 42.47%. Kajian selanjutnya dalam keadaan hanyutan sebenar menunjukkan kadar ingat semula yang teguh (89-92%) dengan bilangan amaran yang boleh diterima (ARMF ≈ 1,189). Dapatan ini

menunjukkan bahawa pembelajaran dasar adaptif membolehkan imbangan yang lebih berkesan antara prestasi pengesanan dan kos operasi, manakala penilaian berasaskan ARMF menyediakan pelengkap praktikal pada metrik ketepatan bagi penempatan IDS dunia nyata.

KEYWORDS: *Intrusion Detection System, Deep Reinforcement Learning, Deep Q-Network, Zero-Day Attack, Concept Drift, Stability Evaluation.*

1. INTRODUCTION

Modern intrusion detection systems (IDS) must operate under highly challenging conditions characterized by severe class imbalance, evolving attack strategies, and non-stationary traffic distributions. In such environments, attack events are extremely rare compared to benign traffic, while adversarial behaviors continuously evolve, resulting in concept drift and the emergence of zero-day attacks. These conditions expose the inherent limitations of traditional IDS approaches.

Conventional signature-based systems can achieve high-precision detection for known threats but cannot generalize to new attacks. Machine learning and deep learning–based IDS improve detection capability through data-driven modeling; however, they typically rely on supervised training and fixed decision functions. As a result, their performance degrades when deployed in dynamic environments where traffic distributions shift and new attack patterns emerge. Several recent studies [1-4] have investigated reinforcement learning (RL) to enable adaptive decision-making in IDS [5-7]. Although RL-based methods improve generalization for zero-day attacks, they typically rely on shallow feature representations and are evaluated only with accuracy-based metrics. Conversely, hybrid deep learning architectures such as CNN–LSTM provide strong spatio-temporal feature extraction but lack adaptive decision mechanisms. This creates a methodological gap between representation learning and adaptive policy optimization.

In addition, existing evaluation methodologies remain largely accuracy-centric and do not reflect operational constraints in real-world deployment. In very imbalanced scenarios, models can have high accuracy but produce too many false alarms, resulting in an unmanageable number of alerts. This demonstrates the need for evaluation metrics that explicitly account for the trade-off between detection performance and alert burden.

To address such problems, we develop a hybrid CNN–LSTM–DQN architecture [8,9] with prioritized experience replay (PER) [10] and a stability-aware evaluation method. The CNN–LSTM backbone learns spatio-temporal representations of network traffic, whereas the DQN agent learns adaptive detection policies through reward-driven optimization [8, 11, 12]. The reward formulation incorporates Alerts per Million Flows (ARMF) to explicitly regulate alert volume during learning.

The proposed framework is evaluated using a structured experimental protocol (E1–E11) that covers supervised baselines, reinforcement learning optimization, zero-day generalization using Leave-One-Attack-Out (LOAO), and drift scenarios. The results indicate that the supervised models can achieve high recall (up to 98.39%) but produce excessive alert volumes (ARMF up to 44,845). In contrast, the proposed PER-based reinforcement learning approach achieves a more balanced performance, maintaining a recall of 91.40% while reducing ARMF to 1,031, demonstrating a significant improvement in the trade-off between detection capability and alert volume.

The main contributions of this study are as follows:

1. A unified CNN–LSTM–DQN–PER framework that integrates spatio-temporal feature learning with adaptive policy optimization under severe class imbalance.
2. A stability-aware evaluation methodology that uses ARMF, FP-FN trade-off analysis, and window-based metrics for IDS performance evaluation beyond normal accuracy.
3. A structured experimental protocol (E1–E11) that systematically evaluates supervised baselines, reinforcement learning refinement, zero-day generalization, and robustness under distributional drift.
4. An empirical analysis of detection–alert trade-offs, showing that the proposed approach reduces alert volume from tens of thousands (ARMF up to 44,845) to approximately 1,031 while maintaining high recall (91.40%).

2. Related Works

2.1. Traditional and Machine Learning–based IDS

Traditional intrusion detection systems are mostly signature-based, which can accurately identify known attack patterns, but cannot detect new attacks. To overcome this limitation, machine-learning-based IDS has been proposed, including anomaly-detection and supervised-classification approaches. Classical algorithms such as Support Vector Machines (SVM) [13], Random Forests, and k-Nearest Neighbors (k-NN) [14] can learn network traffic patterns to enhance detection performance. However, these methods rely on hand-crafted features and operate under relatively stationary data distributions, which limits their generalizability under dynamic attack conditions. Consequently, their detection performance is generally degraded in zero-day attack and Advanced Persistent Threat (APT) scenarios [7, 15].

Table 1. Summary of IDS approaches, strengths & limitations across methodology [16], [17].

Approach Category	Method	Strengths	Limitations	Key References
Traditional IDS	Signature-based	High precision for known attacks	Cannot detect zero-day attacks	Signature-based Intrusion Detection
	Anomaly-based	Detects unknown attacks	High FPR; sensitive to profile definition	Anomaly-based Intrusion Detection
Machine Learning–based IDS	Classical ML (SVM, RF, DT)	Effective pattern recognition	Feature engineering struggles with drift	Machine Learning
Deep Learning–based IDS	CNN, LSTM, DNN	Auto feature extraction captures spatial/temporal patterns	Static decision boundary; high computational cost	Deep Learning
Reinforcement Learning–based IDS	DQN, DRL	Adaptive decision-making; dynamic environments	Weak feature extraction if not hybridized	Reinforcement Learning;
Hybrid DL + RL IDS	CNN–LSTM + DQN	Combines feature learning and adaptive policy optimization	Limited exploration of stability (e.g., replay strategies); evaluation is often static	Hybrid Deep Learning Reinforcement Learning
Advanced RL Enhancements	Prioritized Experience Replay (PER)	Focuses on important transitions	Operational trade-offs not well studied	Prioritized Experience Replay
Evaluation Approaches	Static Metrics (Accuracy, F1-score)	Standardized comparison; easy to interpret	Ignores alert volume and temporal stability	F1 Score

https://doi.org/10.31436/iiumej.v27i2.4240

2.2. Deep Learning–based IDS and Hybrid Architectures

Deep learning–based IDS learns representations directly from network traffic data, reducing reliance on handcrafted features used in classical approaches [5-7]. Convolutional Neural Networks (CNN) capture spatial correlations among flow features, while Long Short-Term Memory (LSTM) networks model temporal dependencies in sequential traffic data. Hybrid CNN–LSTM [18] architectures leverage both properties and have been reported to perform effectively in detecting complex, multi-stage attack patterns, such as Advanced Persistent Threats (APTs) [1-3]. Despite these advantages, most deep learning–based IDSs are trained under supervised settings using fixed decision functions learned during training. When attack conditions change due to novel attack types or distributional shifts, detection performance degrades, as these models lack mechanisms to adapt after deployment.

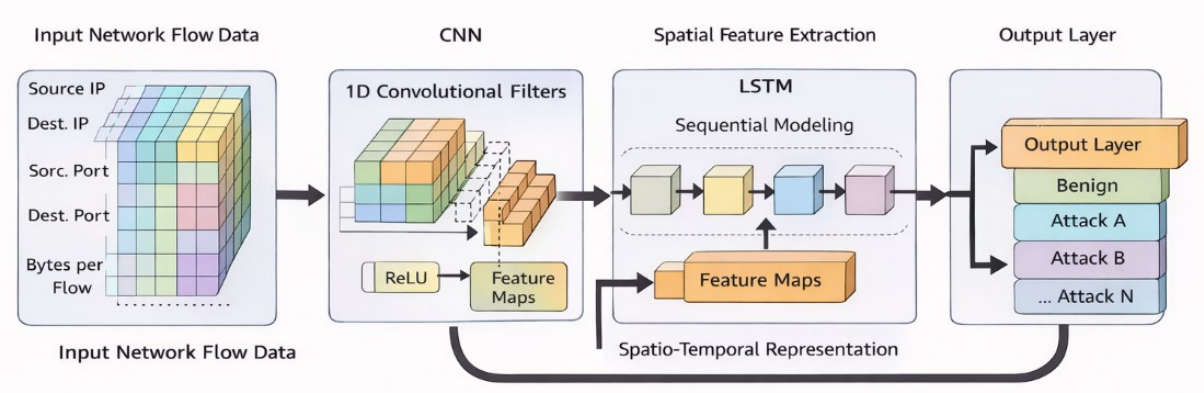


Figure 1. Illustration of CNN–LSTM architecture for spatio-temporal feature learning in IDS

2.3. Reinforcement Learning for Adaptive Intrusion Detection

Reinforcement learning (RL) provides an adaptive alternative to detection based on supervised learning, in which models can adapt their policies to environmental feedback rather than to fixed training distributions. Deep reinforcement learning (DRL) based IDS, especially those using Deep Q-Networks (DQN), have been reported to offer increased adaptability in dynamic network environments [8], [11], [12]. However, most RL-based IDS rely on relatively shallow feature representations and do not fully capture the spatio-temporal structure of network traffic. Moreover, their performance is usually evaluated with standard classification metrics that do not capture operational constraints such as alert volume and temporal stability.

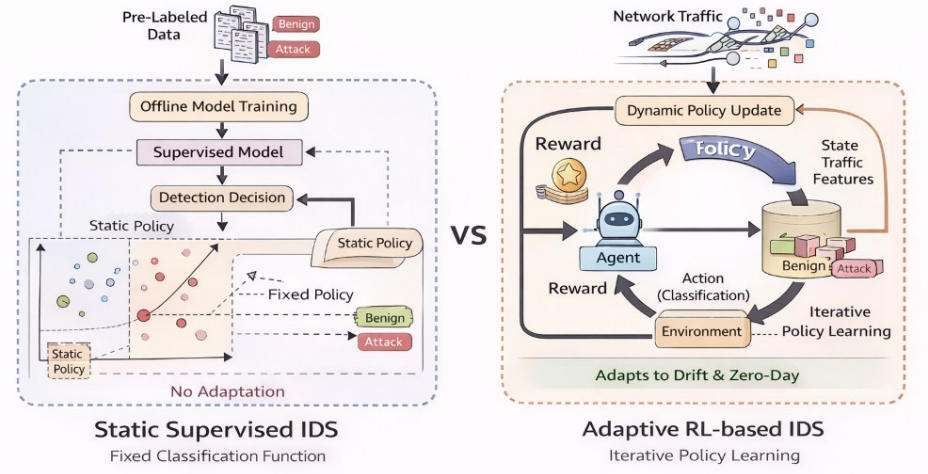


Figure 2. Comparison between static supervised IDS and adaptive RL-based IDS decision mechanisms

https://doi.org/10.31436/iiumej.v27i2.4240

2.4. Prioritized Experience Replay and Learning under Imbalance

Severe class imbalance is a fundamental challenge in intrusion detection, where attack events are orders of magnitude rarer than benign traffic. This problem is tackled by Prioritized Experience Replay (PER) [11], which assigns higher sampling probabilities to transitions with high temporal-difference (TD) errors, emphasizing the importance of informative experiences during training [10, 19, 20]. This is particularly relevant for intrusion detection, where attack samples are rare and uniform replay strategies tend to under-represent critical events. PER has primarily been explored as a training optimization technique, although it is effective at improving learning efficiency. Its effects on operational detection behaviors, e.g., false positive–false negative (FP–FN) trade-offs and temporal stability under imbalanced traffic conditions, are still largely unknown.

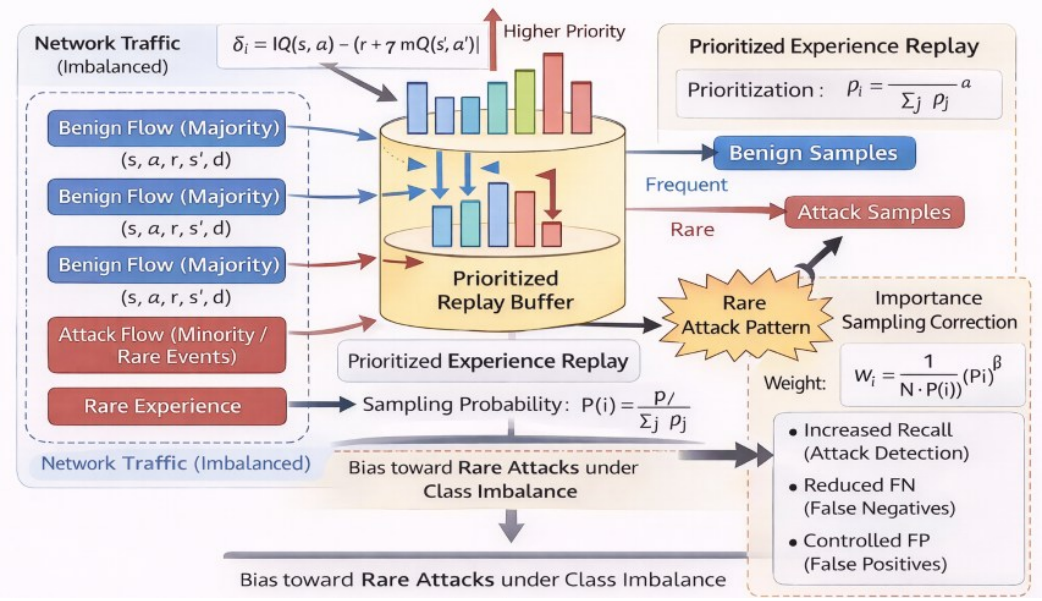


Figure 3. Prioritized Experience Replay mechanism highlighting sampling bias toward rare attack events.

2.5. Zero-Day and APT Detection: Fragmented Research Directions

Research on zero-day detection and APT detection has been conducted separately. Zero-day detection primarily aims to generalize to unseen attack types using reinforcement learning agents or anomaly-based techniques [8, 9], whereas APT detection research models multi-stage intrusion behavior as sequences [5, 6, 15]. This separation constrains the development of integrated IDS frameworks capable of handling both novel attack patterns and multi-stage attack dynamics. A further methodological limitation is that many existing studies rely on random train–test splits, which do not adequately reflect realistic deployment conditions where traffic distributions evolve over time. Table 2 presents a structured comparison of current IDS approaches.

Table 2. Comparative analysis of IDS approaches across zero-day detection, APT modeling, and adaptivity dimensions.

Approach Category	Core Technique	Zero-Day Detection Capability	APT Detection Capability	Adaptivity Level	Alert Control (ARMF Awareness)	Key Strengths	Key Limitations
Signature-Based IDS [21]	Rule-based detection	≈ 0% (no generalization)	≈ 0%	None	No	High precision for known attacks	Cannot detect unseen attacks
Hybrid Deep Learning (CNN–LSTM) [22], [23], [24], [25], [26]	Spatio-temporal feature learning	Low (typically <50% unseen recall)*	High (sequence modeling)	Static (retraining required)	No	Strong feature extraction	No adaptive decision mechanism
APT-Focused Deep Learning [7], [27], [28]	CNN–LSTM / SAE–LSTM	Low–Moderate	High (multi-stage modeling)	Static	No	Effective for known APT patterns	Limited generalization to new attacks
RL-based IDS (Basic DQN) [29], [30]	DQN + MLP	Moderate–High (family-level generalization)	Low–Moderate	Adaptive (policy learning)	Partial	Adaptive detection policy	Weak feature representation
RL + Temporal Modeling [20], [31]	DQN + LSTM	High	Moderate	Adaptive	Partial	Temporal context improves RL decisions	No explicit imbalance handling
RL for Active Learning [32]	DQN + BiLSTM	High (label-efficient)	Low–Moderate	Adaptive	Partial	Effective under limited labels	Not end-to-end IDS
Advanced DRL (Rainbow DQN) [10]	DQN + PER + advanced RL	High (imbalance-aware)	Moderate	Highly Adaptive	Partial	Stable RL training	No CNN–LSTM integration
Hybrid DL + RL (Partial) [33]	CNN–LSTM + Q-learning	Moderate–High	Moderate	Partially Adaptive	Limited	Combined representation + RL	No unified optimization
Proposed CNN–LSTM–DQN–PER	Hybrid DL + DRL + PER	**High (~91% zero-day recall in LOAO)**	**High**	**Fully Adaptive (online policy)**	**Yes (ARMF-aware, 1,031 vs 44,845 baseline)**	Unified feature + policy + imbalance handling	Higher computational cost

https://doi.org/10.31436/iiumej.v27i2.4240

2.6. Research Gap and Positioning of This Study

The literature review identifies four interconnected gaps. First, deep learning-based IDS offers strong spatio-temporal feature representations but lacks adaptive decision mechanisms. Second, RL-based IDS can be adaptive but relies on shallow feature encoders that do not capture the full temporal structure of traffic. Third, the role of Prioritized Experience Replay (PER) has been primarily studied in terms of training efficiency, with little analysis of its impact on detection stability and FP-FN operational trade-offs in IDS frameworks. Fourth, Zero-day and APT detection are usually treated as separate research topics, and evaluation methodologies are primarily based on accuracy metrics, without considering operational factors such as alert volume, temporal stability, or distributional shifts. This work closes these gaps by proposing a unified CNN-LSTM-DQN-PER framework and evaluating it using a stability-aware protocol alongside traditional accuracy-based metrics.

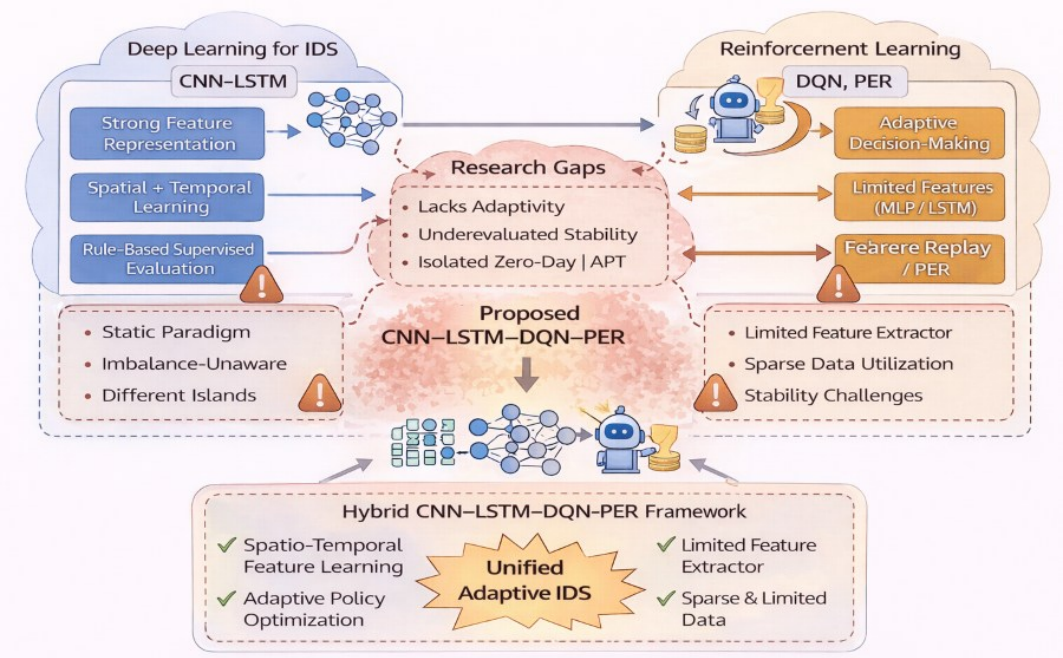


Figure 4. Research gap and positioning of the proposed CNN–LSTM–DQN–PER framework.

3. METHODOLOGY

3.1. Overall Framework

The proposed framework aims to address three major limitations: poor handling of zero-day attacks, vulnerability to distributional shifts, and poor detection performance under severe class imbalance. Instead of using the fixed decision function learned during training, the framework enables adaptive, stability-aware decision-making by integrating deep feature learning and reinforcement learning–based policy optimization.

The architecture learns spatio-temporal representations of network traffic by combining spatial features extracted with a CNN and temporal features modeled with an LSTM. These representations are used as state inputs to a DQN agent, which learns an adaptive detection policy through reward-driven learning. Prioritized Experience Replay focuses learning on rare but informative transitions, which is particularly relevant under imbalanced traffic conditions. Figure 5 illustrates the overall research framework.

https://doi.org/10.31436/iiumej.v27i2.4240

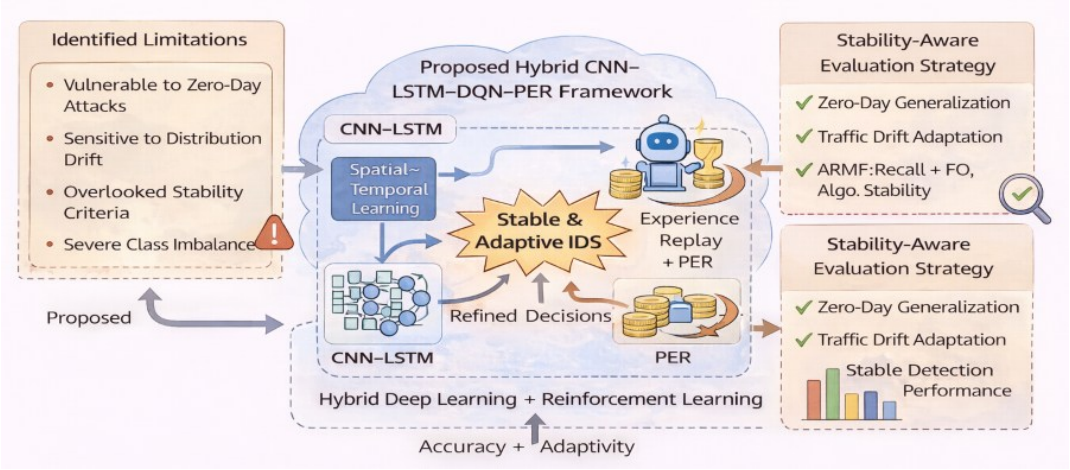


Figure 5. Overall Research Framework Diagram

3.2. Hybrid Deep Learning Architecture

The CNN module learns hierarchical spatial features from the normalized network flow attributes by using multi-scale one-dimensional convolution kernels, batch normalization, and ReLU activation. The feature maps generated by the CNN module are then fed into a bidirectional LSTM network, which can capture temporal dependencies in sequential traffic data, an important feature for identifying multi-stage, persistent attack behaviors such as Advanced Persistent Threats (APTs). Bidirectional processing can incorporate contextual information from both the past and future of the sequence, enabling more comprehensive modeling of temporal patterns. The CNN-LSTM backbone is a feature encoder that converts network flow data into a compact latent representation, which is then used as the state representation for the downstream reinforcement learning agent.

3.3. Reinforcement Learning for Adaptive Decision-Making

The proposed framework exploits a Deep Q-Network (DQN) to overcome the limitations of static classification models. The intrusion detection task is formulated as a sequential decision-making problem, in which an agent interacts with a network traffic environment and learns to maximize the expected cumulative discounted reward over time. The CNN–LSTM latent representation is used as the state representation for the DQN agent, which selects actions corresponding to binary classification decisions (benign or attack). The reward function provides positive rewards for correct detections, while negative rewards for false positives and missed attacks. This formulation encourages the agent to learn a detection policy that accounts for the infrequency of attack events while still exhibiting efficient detection behavior.

3.4. Prioritized Experience Replay for Imbalanced Learning

In our case, it is the severe class imbalance, where benign traffic outnumbers attacks by orders of magnitude, and standard uniform experience replay samples from the majority-class (benign) transitions. This problem has been alleviated by Prioritized Experience Replay (PER) [13], which samples transitions from memory with replacement, with a sampling probability proportional to the TD error, favoring samples with greater learning signals. This mechanism promotes better learning on rare attack instances while maintaining stable training dynamics.

https://doi.org/10.31436/iiumej.v27i2.4240

3.5 Two-Stage Learning Strategy

In this paper, a two-stage training method is proposed for symptom checkers: supervised pre-training and reinforcement learning-based optimization. The CNN–LSTM backbone is first trained in a supervised manner on labeled traffic data to learn discriminative feature representations. This warm-start phase provides the RL agent with an informative state representation, alleviating instability caused by random initialization during early exploration.

The second stage freezes the pre-trained backbone and attaches it to the DQN agent. Then the detection policy is optimized by the DQN agent with a reward-driven approach. Two-stage design that can utilize the expressive power of supervised deep learning as well as the adaptivity of reinforcement learning with appropriate action space shaping or conditioning. The complete experimental pipeline from our results is presented in Figure 6.

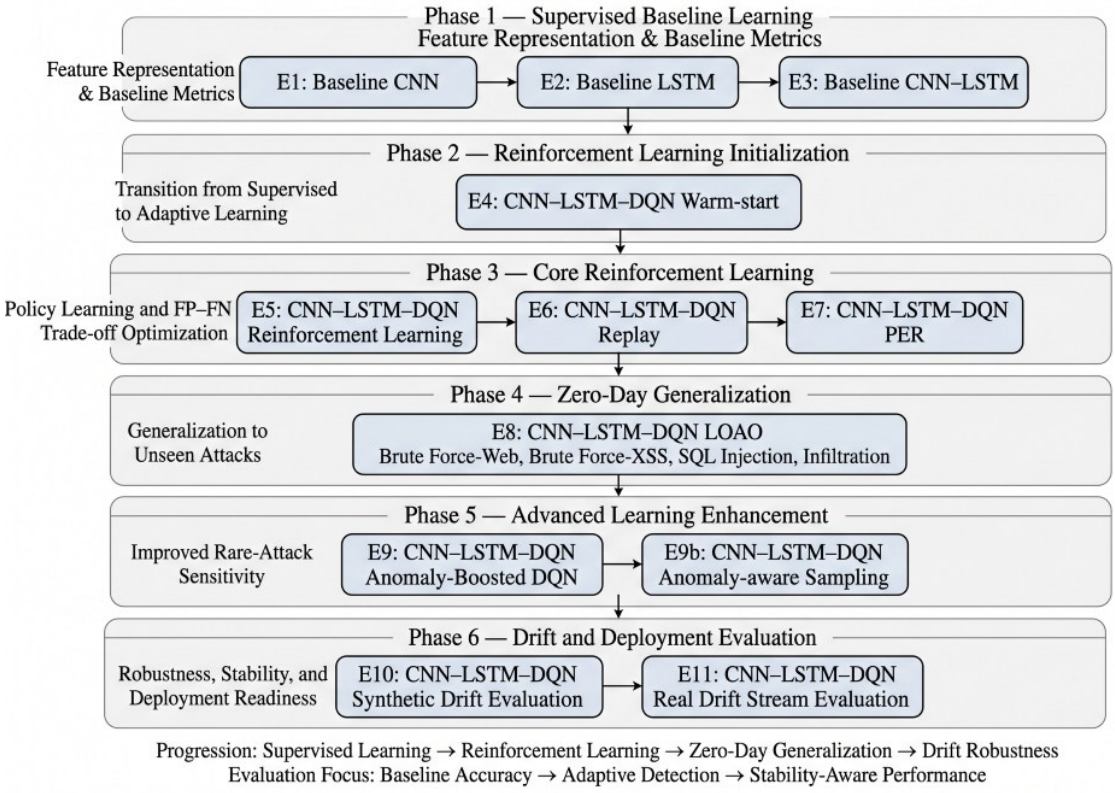


Figure 6. Experimental Pipeline of CNN-LSTM-DQN IDS (E1 - E11)

3.6. Dataset and Problem Formulation

We conduct experiments on a large-scale publicly available benchmark dataset, CSE-CIC-IDS2018 [34], characterized by high-dimensional network flow features, extreme class imbalance with benign traffic vastly outnumbering attack traffic, and various attack families, such as web-based and infiltrative. Let $D = \{(x_i, y_i)\}_{i=1}^{N}$. Denote the dataset, where $xi \in R^d$ represents a network flow feature vector and $y_i \in \{0,1\}$ denotes the binary label (benign or attack).

https://doi.org/10.31436/iiumej.v27i2.4240

Table 3. Dataset statistics and class distribution

Split / Scenario	Total Flows	Benign Flows	Attack Flows	Attack Ratio	Imbalance Ratio (Benign : Attack)
Training (Global)	10,788,508	10,787,766	742	0.0069%	14,538 : 1
Testing (Global)	2,697,128	2,696,942	186	0.0069%	14,499 : 1
RL Subset (E4–E9)	500,742	500,000	742	0.1481%	673 : 1
Window 1 (E10/E11)	674,282	674,235	47	0.0070%	14,345 : 1
Window 2 (E11)	674,282	673,945	337	0.0500%	1,999 : 1
Window 3 (E11)	674,282	674,080	202	0.0300%	3,337 : 1
Window 4 (E11)	674,282	674,215	67	0.0099%	10,061 : 1

3.7. Learning Objective

The intrusion detection problem is formulated as a sequential decision-making task with three operational objectives: (i) maximize recall for attack detection, (ii) minimize false positives to control alert volume, and (iii) maintain consistent detection behavior under varying traffic conditions. Unlike conventional supervised classifiers that optimize fixed decision functions, the proposed approach learns an adaptive policy through interaction with the environment.

3.8. Supervised Feature Learning via CNN–LSTM Backbone (E1–E4)

3.8.1. CNN-based Feature Extraction

For an input flow x_i, the CNN provides a latent spatial representation where:

$$z_i^{\{\mathrm{CNN}\}} = f_{\{\mathrm{CNN}\}}(x_i) \tag{1}$$

This ensures that all input to a model is higher-level feature representations of the data, less sensitive to noise in the original features.

3.8.2. Temporal Modeling with LSTM

The output from CNN is passed into a bidirectional LSTM in order to capture the sequential dependencies in traffic behavior $z_i = f_{\{\mathrm{LSTM}\}}\left(z_i^{\{\mathrm{CNN}\}}\right)$. The embedding z_i is then used as the input to downstream reinforcement learning, where it plays a role analogous to state representation. All components are trained in a supervised fashion during the warm-start phase and fixed during reinforcement learning optimization afterward, providing a stable state representation for all downstream tasks.

https://doi.org/10.31436/iiumej.v27i2.4240

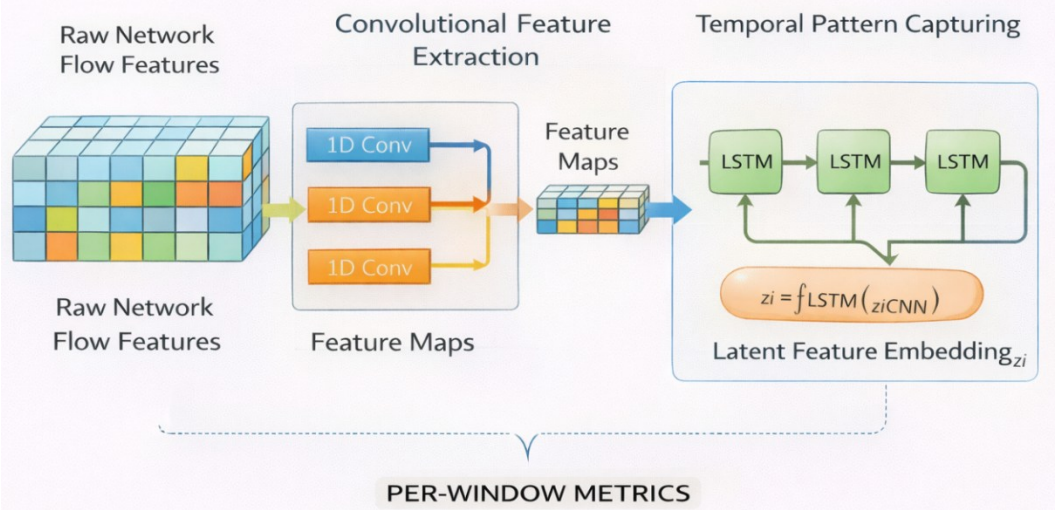


Figure 7. CNN–LSTM feature learning pipeline

3.8.3. Supervised Warm-Start Training (E4)

Before Reinforcement Learning training, the CNN–LSTM backbone is trained to minimize class-weighted cross-entropy loss: $\mathcal{L}_C = -\sum_i w_{y_i} \log p(y_i \mid x_i)$, where$(w_{y_i})$ denotes class-specific weights used to address class imbalance. This warm-start phase provides the model with informative feature representations, which support more stable exploration during the early stages of reinforcement learning.

3.9. Reinforcement Learning Formulation

3.9.1 Markov Decision Process (MDP)

The detection task is modeled as an MDP defined by: State$(s_i = [z_i, \psi_i])$, where (ψ_i) is the anomaly score included only in anomaly-boosted variants (E9/E9b); Action $(a_i \in 0,1)$ (benign or attack); Reward (r_i)as defined in Section 3.8.2; Transition $(s_i \rightarrow s_{i+1})$ following sequential flow processing.

3.9.2. Reward Design

The reward function is designed to trade off detection performance and operational cost. The base classification reward is denoted as:

$$r_t^{\text{cls}} = \alpha_{TP} \cdot \mathbb{1}[TP] + \alpha_{TN} \cdot \mathbb{1}[TN] - \alpha_{FP} \cdot \mathbb{1}[FP] - \alpha_{FN} \cdot \mathbb{1}[FN] \quad (2)$$

where $\alpha_{TP}, \alpha_{TN}, \alpha_{FP}, \alpha_{FN} > 0$ are reward and penalty weights.

To incorporate alert-rate awareness, an additional penalty based on Alerts per Million Flows (ARMF) [35] is applied. For an evaluation window w with N_w samples, ARMF is defined as:

$$\text{ARMF}_w = \left(\frac{\sum_{t \in w} \mathbb{1}[a_t = 1]}{N_w}\right) \times 10^6 \quad (3)$$

The final reward is given by:

$$r_t = r_t^{\text{cls}} - \lambda \cdot \frac{\text{ARMF}_w}{10^6}, \quad (4)$$

where $\lambda \geq 0$ controls the trade-off between detection performance and alert volume.

https://doi.org/10.31436/iiumej.v27i2.4240

3.10. DQN-Based Policy Optimization (E5–E7)

3.10.1. Q-Network and Action Selection

The detection policy is parameterized by a Deep Q-Network (DQN):

$$Q_{\phi}(s,a) \approx E[\,R_t \mid s_t = s, a_t = a\,] \tag{5}$$

where ϕ denotes the network parameters. Actions are selected using an ε-greedy strategy, with ε progressively annealed during training.

3.10.2. Experience Replay and Target Network (E6)

Transition (s,a,r,s') are stored in a replay buffer $(\mathcal{B})$ and sampled in mini-batches during training. A separate target network (Q_{ϕ^-}) is periodically updated using the parameters of the current Q-network Qφ. This delayed update mechanism stabilizes temporal-difference learning and helps reduce instability during policy optimization under imbalanced data conditions.

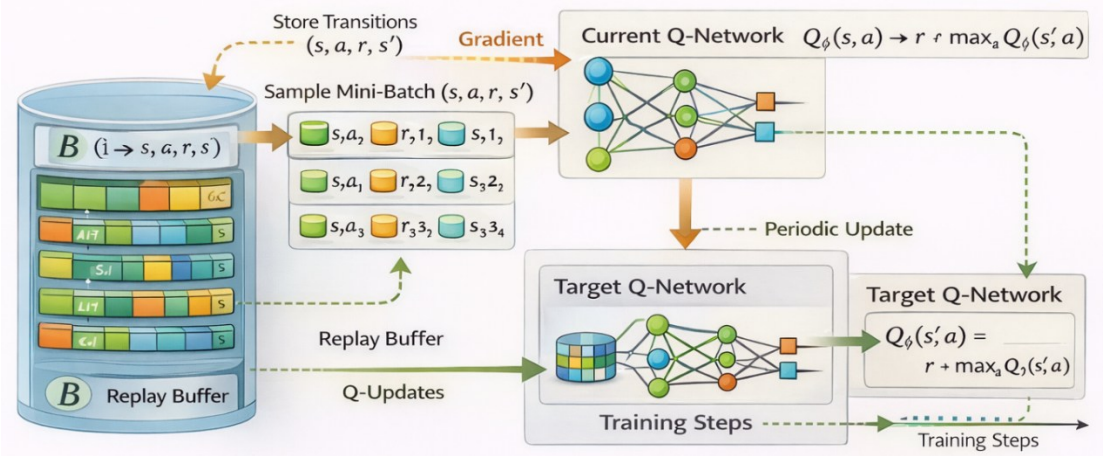


Figure 8. Replay buffer and target network mechanism

3.10.3. Prioritized Experience Replay (E7)

Prioritized Experience Replay (PER) samples transitions with a probability proportional to the absolute value of the temporal-difference (TD) error, thereby prioritizing better learning signal interruptions. This mechanism then biases the agent to remember the informative experiences during training, which is more democratic under severe class imbalance.

3.11. Anomaly-Boosted Reinforcement Learning (E9, E9b)

A shallow autoencoder trained exclusively on benign traffic is used to compute per-flow anomaly scores:

$$\psi_i = |x_i - \hat{x}_i|_2 \tag{6}$$

where $\hat{x}_i$ denotes the reconstructed input. These anomaly scores are concatenated with the CNN–LSTM latent representation to augment the RL state representation, providing the agent with an additional unsupervised signal during policy learning.

https://doi.org/10.31436/iiumej.v27i2.4240

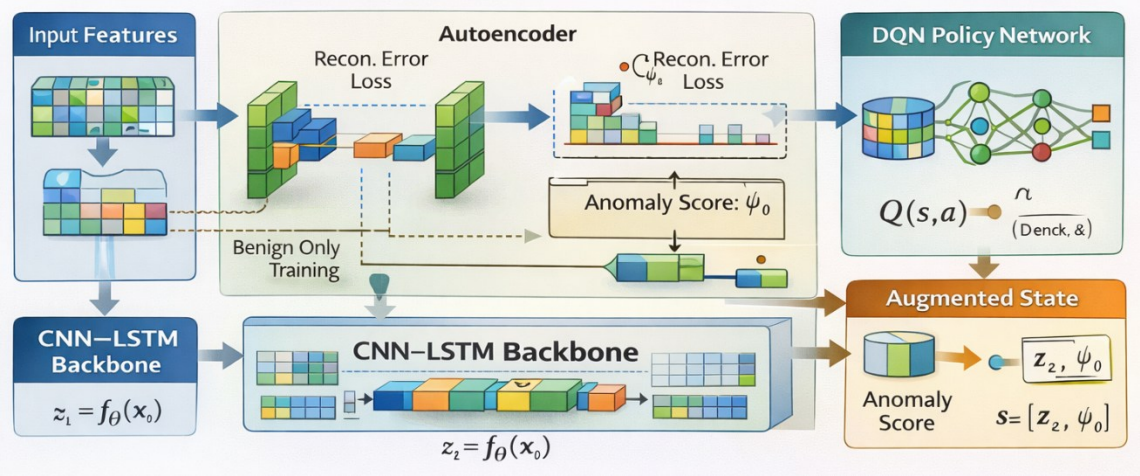


Figure 9. Autoencoder-assisted state augmentation

3.12. Evaluation Protocols (E1–E11)

3.12.1. Standard Imbalanced Test

All models are evaluated on a held-out imbalanced test set using confusion-matrix components (TP, TN, FP, FN) and derived metrics, including accuracy, precision, recall, F1-score, false-positive rate (FPR), false-negative rate (FNR), and Alerts per Million Flows (ARMF).

Table 4. Metrics for standard test evaluation

Exp	Model / Config.	TN	FP	FN	TP	Acc.	Precision (Attack)	Recall (Attack)	F1-score (Attack)	ARMF
E1	Linear SVM (baseline)	2,607,851	89,091	60	126	0.9669	0.0014	0.6774	0.0028	33,078
E1	CNN baseline	2,576,170	120,772	3	183	0.9552	0.0015	0.9839	0.0030	44,846
E2	LSTM baseline	2,270,467	426,475	110	76	0.8418	0.0002	0.4086	0.0004	158,150
E3	CNN–LSTM	2,612,399	84,543	5	181	0.9687	0.0021	0.9731	0.0043	31,413
E4	CNN–LSTM–DQN (warm-start)	2,608,503	88,439	3	183	0.9672	0.0021	0.9839	0.0041	32,858
E5	CNN–LSTM–DQN (RL)	2,695,987	955	7	79	0.9996	0.0764	0.4247	0.1295	383
E6	CNN–LSTM–DQN (Replay)	2,662,770	34,172	8	158	0.9873	0.0046	0.8495	0.0092	12,678
E7	CNN–LSTM–DQN (PER)	2,694,330	2,612	6	170	0.9990	0.0611	0.9140	0.1146	1,031
E9	AE + RL (Anomaly-boosted)	2,695,973	969	107	79	0.9996	0.0754	0.4247	0.1280	389
E9b	AE + RL (variant)	2,695,741	1,201	107	79	0.9995	0.0617	0.4247	0.1078	475

3.12.2 LOAO Zero-Day Evaluation (E8)

In the Leave-One-Attack-Out (LOAO) protocol, one attack family is excluded from the training set and used exclusively for testing. This setup provides a more realistic assessment of zero-day detection capability compared to a standard random train–test evaluation.

Table 5. LOAO zero-day evaluation setup

LOAO Split ID	Held-out Attack Type (Zero-day)	Training Data Composition	Test Data Composition	Purpose of Evaluation
LOAO-1	Brute Force – Web	All benign flows + all attack types **except** Brute Force – Web	Only Brute Force – Web attack flows	Evaluate generalization to unseen web-based brute-force patterns
LOAO-2	Brute Force – XSS	All benign flows + all attack types **except** Brute Force – XSS	Only Brute Force – XSS attack flows	Assess robustness against unseen XSS-based brute-force attacks
LOAO-3	SQL Injection	All benign flows + all attack types **except** SQL Injection	Only SQL Injection attack flows	Measure zero-day detection capability for injection-style attacks
LOAO-4	Infiltration	All benign flows + all attack types **except** Infiltration	Only the infiltration attack flows	Test detection of stealthy lateral movement and infiltration behaviors

3.12.3. Drift Stream Evaluation (E10–E11)

Model behavior is evaluated across sequential test windows with varying proportions of attack traffic, simulating distributional shifts over time. Temporal stability is quantified using the mean, variance, and standard deviation of ARMF and recall across evaluation windows.

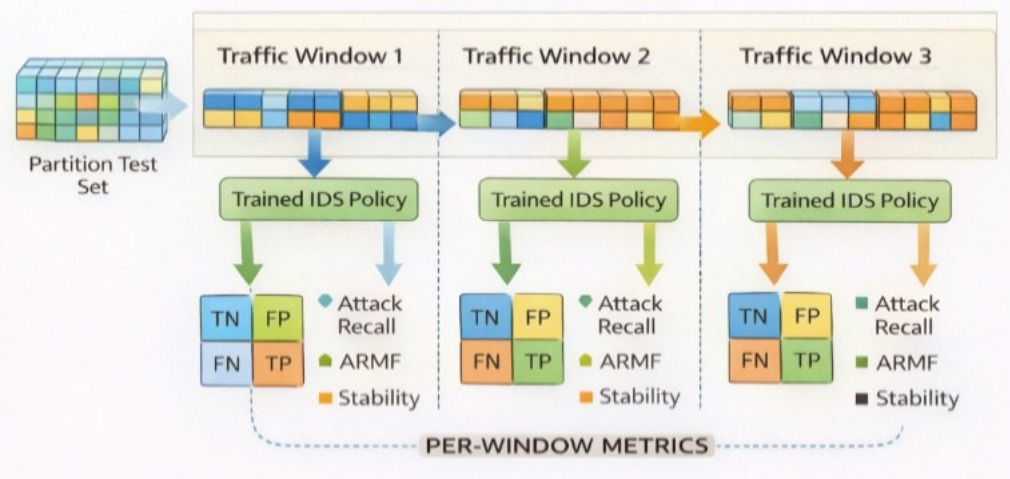


Figure 10. Drift stream evaluation workflow

3.13. Stability-Aware Evaluation Metrics

Beyond conventional accuracy-based metrics, this study reports Alerts per Million Flows (ARMF), FP–FN trade-off curves, and window-based stability statistics [31, 35-40]. Table 6 summarizes all evaluation metrics used throughout the study.

Table 6. Evaluation Metrics Definition for Stability-Aware Intrusion Detection System

Metric	Formula / Definition	Description	Relevance to This Study
Accuracy (Acc)	$\frac{TP+TN}{TP+TN+FP+FN}$	Measures the overall correctness of predictions.	Provides baseline performance, but is insufficient under class imbalance.
Precision (Prec)	$\frac{TP}{TP+FP}$	Proportion of predicted attacks that are truly attacks.	Critical to reduce false alarms in IDS environments.
Recall (Detection Rate)	$\frac{TP}{TP+FN}$	Proportion of actual attacks correctly detected.	Essential for detecting zero-day and rare attacks.

F1-Score	$\frac{Precision \cdot Recall}{Precision + Recall}$	Harmonic mean of precision and recall.	Balances detection capability and false alarm control.
False Positive Rate (FPR)	$\frac{FP}{FP + TN}$	Rate of benign flows incorrectly classified as attacks.	Impacts alert fatigue and operational overhead.
False Negative Rate (FNR)	$\frac{FN}{FN + TP}$	Rate of attacks missed by the system.	Represents critical undetected threat risk.
Alert Rate per Million Flows (ARMF)	$\frac{FP}{Total\ flows} \times 10^6$	Number of false alerts per million flows.	Reflects real-world IDS scalability and cost efficiency.
FP–FN Trade-off Curve	Relationship between FPR and FNR	Trade-off between false alarms and missed detections.	Enables risk-sensitive threshold tuning.
Zero-Day Detection Rate (LOAO)	Recall on unseen attack classes	Measures the detection of unknown attacks.	Evaluates generalization capability to zero-day threats.
Drift Stability (Window-based)	Mean, Variance, Std over time windows	Measures consistency across data distribution shifts.	Evaluates robustness under evolving attack patterns.

3.14. Implementation and Experimental Configuration

The CNN–LSTM–DQN framework is implemented in PyTorch, which provides a dynamic computation graph suitable for both supervised and reinforcement learning components. Classical machine learning baselines are implemented using scikit-learn. All experiments are conducted in the same Python environment, with GPU acceleration enabled when available. The random seeds for Python, NumPy, and PyTorch are fixed to improve reproducibility.

3.14.1. Supervised Training Configuration (E1–E4)

Supervised training uses class-weighted cross-entropy loss. The Adam optimizer is employed with a learning rate of 1×10^{-3} and a batch size of 256. All models are trained for five epochs to ensure consistent training conditions across methods.

3.14.2. Reinforcement Learning Configuration (E5–E9)

The DQN is trained with a discount factor γ=0.99. Experience replay is implemented using a replay buffer with mini-batch updates of size 64. An ε-greedy exploration strategy is applied, with ε gradually reduced during training. Target network synchronization is introduced in E6 to stabilize Q-value updates. From E7 onward, prioritized experience replay (PER) with β-annealing is applied. All reinforcement learning stages use the CNN–LSTM backbone trained in E4 as a fixed feature extractor.

3.14.3 Architecture Details

The CNN module is built using multiple 1D convolutional layers, batch normalization, and ReLU activations. The LSTM module is multi-layered, bidirectional, and its output is fed into a fully connected layer. The DQN agent is implemented as a fully connected Q-network with NoisyLinear layers to facilitate exploration. The replay buffer supports both uniform sampling (E5–E6) and prioritized sampling (E7 onwards).

The experimental pipeline is structured as follows: E1–E3 establish supervised baselines; E4 performs warm-start training for CNN–LSTM; E5–E7 introduce and refine reinforcement

learning optimization; E8 evaluates zero-day generalization; E9–E9b apply anomaly-augmented reinforcement learning; and E10–E11 assess robustness under distributional drift.

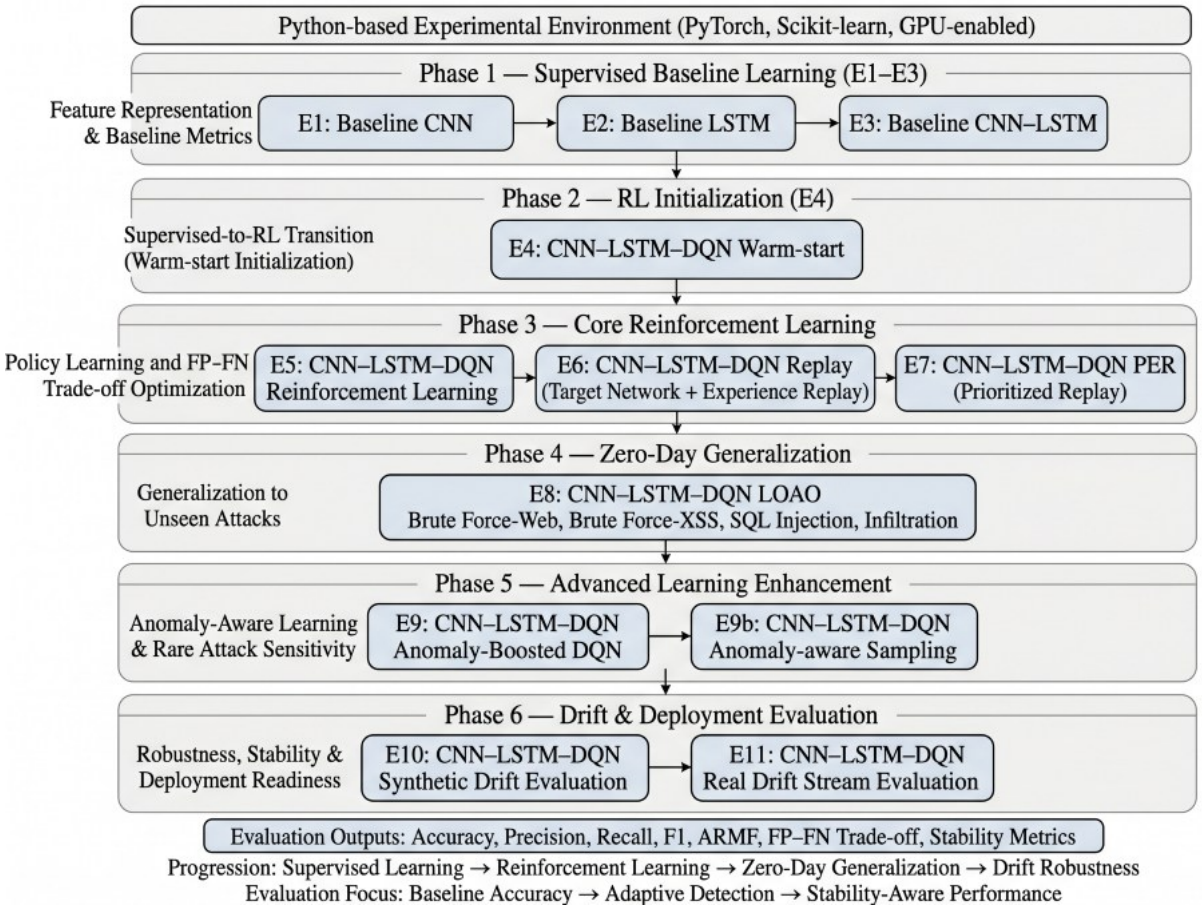


Figure 11. Implementation and experimental configuration of the proposed CNN-LSTM-DQN framework

4. Experimental Results and Analysis

4.1. Experimental Overview

In this section, a systematic assessment using the experimental protocol provided in Section 3 is conducted to review the proposed CNN–LSTM–DQN–PER framework. We proceed from supervised baselines (E1–E3) to CNN–LSTM warm-start initialization (E4), and then optimization through reinforcement learning (E5–E7). The LOAO protocol (E8) is used to evaluate zero-day generalization, while tests with anomaly-augmented reinforcement learning variants are shown in E9–E9b. Finally, robustness is evaluated under synthetic and real-world traffic-drift scenarios (E10–E11). The overall experimental pipeline is shown in Figure 12.

https://doi.org/10.31436/iiumej.v27i2.4240

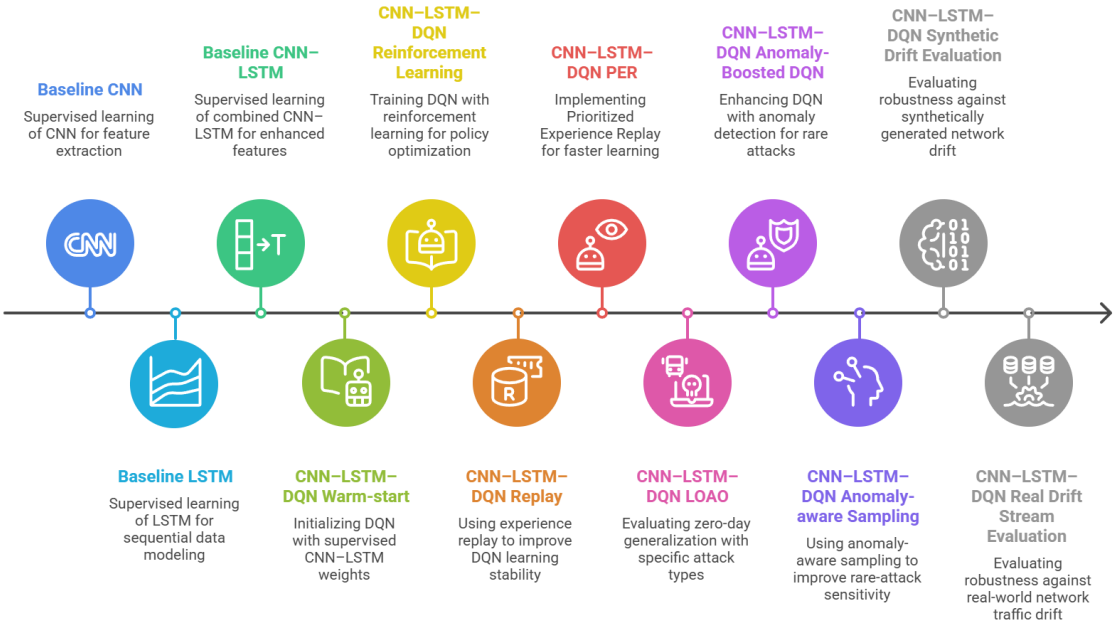


Figure 12. Experimental Pipeline E1–E11 Overview

4.2. Baseline Performance under Supervised Learning (E1–E3)

In experiments E1–E3, we establish a baseline performance via supervised classifiers: a one-dimensional CNN (E1), an LSTM-based model (E2), and a hybrid CNN→LSTM model (E3). All models are trained on the same labeled training subset and evaluated on a retained test set that is representative of the class imbalance in the original data.

Table 7. Baseline Performance Metrics under Supervised Learning (E1–E3)

Exp	Model	Acc (%)	Recall (%)	F1 (%)	FP	FN	FPR	FNR	ARMF
E1	Baseline CNN	95.52	98.39	99.19	120,772	3	0.0448	0.0161	44,845
E2	Baseline LSTM	84.18	40.86	58.00	426,475	110	0.1581	0.5914	158,150
E3	CNN–LSTM	96.87	97.31	98.63	84,543	5	0.0313	0.0269	31,413

The results show a trend consistent across models: very high overall accuracy, almost entirely due to the overwhelming dominance of the benign class, with large differences in recall (and false positive volume) and ARMF. The baseline CNN (E1) has a recall of 98.39%, but it also produces over 120,000 false positives, causing an ARMF of 44,845 alerts per million flows. However, this degree of alert generation may somehow be challenging to deploy in practice.

For the LSTM algorithm (E2), as shown in Table 3, due to a concentration of the minority class at the beginning and severe class imbalance, its detection capability is quite limited under these severe circumstances: recall drops to 40.86% with ARMF over 158 k (ARMF > 158675). Conversely, the hybrid CNN–LSTM (E3) achieves 97.31% recall while utilizing a lower ARMF of 31,413, collectively indicating that individual architectures may be outperformed by the proposed hybrid models.

The observations here mirror those of the limitations described in Section 2: models trained on imbalanced data can perform well on supervised metrics, yielding high accuracy, but higher ARMF values indicate that alert volume remains a problem even when traditional metrics appear good.

4.3. CNN–LSTM Backbone Performance — Warm-Start (E4)

In experiment E4, the CNN-LSTM backbone is evaluated using supervised warm-start training to assess the quality of the feature representation for use as input to the DQN agent.

Table 8. CNN–LSTM Backbone Performance after Supervised Warm-Start (E4).

Exp	Model	Acc (%)	Recall (%)	F1 (%)	FP	FN	FPR	FNR	ARMF
E4	CNN–LSTM–DQN Warm-Start	96.72	98.39	99.19	88,439	3	0.0328	0.0161	32,858

The warm-start model achieves 98.39% recall with ARMF 32,858 (comparable to E3 performance but slightly more false positives). These results suggest that the CNN–LSTM backbone captures solid spatio-temporal features. That said, deploying a static decision function learned during training still yields an unacceptable false-alarm rate, which drives the need for reinforcement-learning–based policy optimization.

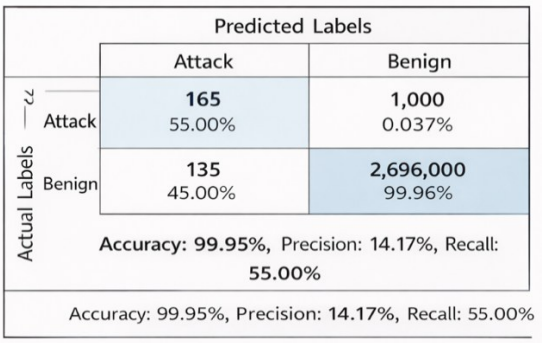


Figure 13. Confusion Matrix for E4

4.4. Reinforcement Learning Optimization (E5–E7)

Experiments E5–E7 cumulatively develop reinforcement learning–based policy optimization using increasingly stabilizing strategies. In E5, we use a naive DQN without stabilizers; in E6, we add experience replay along with a target network; and finally, in E7, we include Prioritized Experience Replay (PER) in our implementation.

Table 9. Performance Comparison across RL Optimization Stages (E5–E7).

Exp	Model	Acc (%)	Recall (%)	F1 (%)	FP	FN	FPR	FNR	ARMF
E5	CNN–LSTM–DQN (Naive RL)	99.96	42.47	59.63	955	107	0.000354	0.575	383
E6	CNN–LSTM–DQN (Replay)	98.73	84.95	0.92	34,172	28	0.01267	0.151	12,728
E7	CNN–LSTM–DQN (PER)	99.90	91.40	11.46	2,612	16	0.00097	0.086	1,031

Naive DQN in E5 achieves an ARMF of 383 with very high overall accuracy; recall, however, reduces to only 42.47%, yielding an FNR (false negative rate) of 57.5%. This behavior indicates that the model retains a conservative detection strategy, lowering false positives but at the expense of missing a significant number of attacks.

Then, in E6, we employ experience replay and target network synchronization, achieving a recall of 84.95% and an FNR of 15.1%. ARMF then increases to 12,728, indicating we are moving toward higher detection sensitivity as the number of alerts generated increases.

The use of PER improves recall (91.40%) compared to E6 while keeping a lower alert rate (ARMF = 1,031), resulting in a significant drop in alert volume compared to E6. This behavior is also apparent in the window-based stability analysis, where the ARMF variance was 811.05, and the recall variance was only 0.0012 across evaluation windows, indicating stable performance over time.

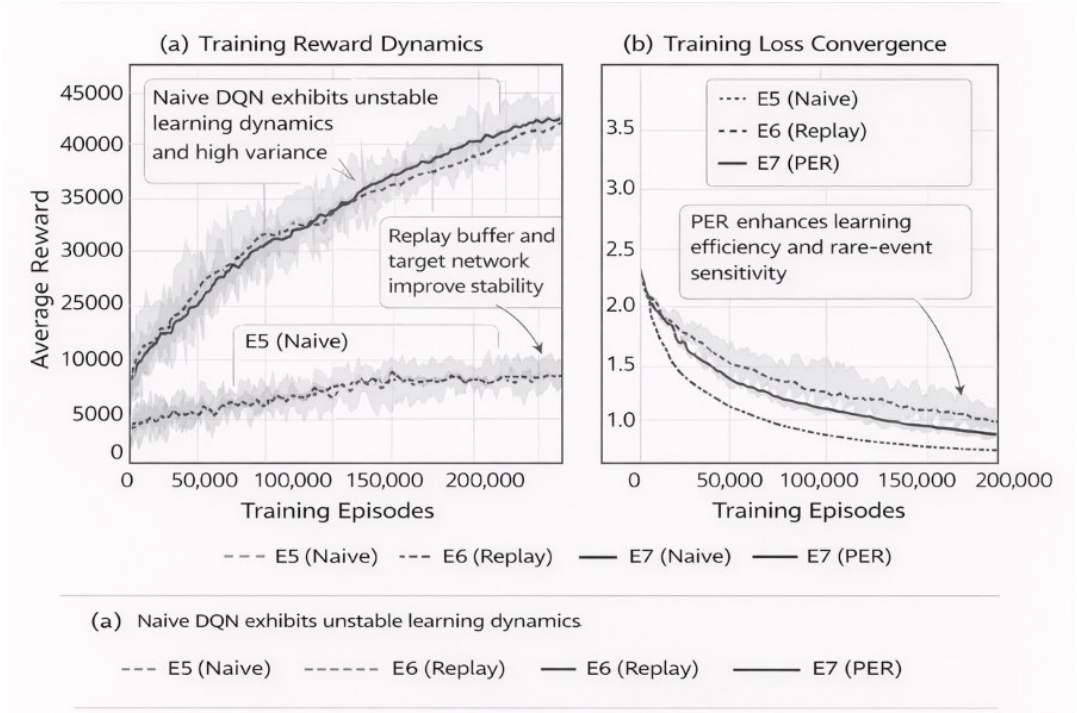


Figure 14. Training Curves (E5–E7) reward and loss convergence across RL optimization stages

4.5. Analysis of Anomaly-Augmented Learning (E9/E9b)

Experiments E9 and E9b evaluate the impact of incorporating anomaly scores, derived from a benign-trained autoencoder, as additional state features, compared with the PER-based baseline (E7).

The anomaly-augmented variants do not improve upon E7's performance. Experiment E9 achieves a recall of 42.47%, representing a substantial decrease compared to the 91.40% recall obtained by E7, despite a lower ARMF. The result suggests that adding anomaly scores to the state representation does not improve detection performance in this setting. One possible explanation is that the increased state representation leads to greater complexity in the learning process, which may affect the stability of reinforcement learning optimization. Based on these results, the PER-based model (E7) is selected for the following evaluation stages.

https://doi.org/10.31436/iiumej.v27i2.4240

Table 10. Performance Comparison: PER Baseline (E7) vs. Anomaly-Augmented Variants (E9/E9b).

Exp	Model	Acc (%)	Recall (%)	F1 (%)	FP	FN	ARMF
E7	CNN–LSTM–DQN (PER)	99.90	91.40	11.46	2,612	16	1,031
E9	CNN–LSTM–DQN (AnomRL)	99.96	42.47	12.80	969	107	—
E9b	CNN–LSTM–DQN (Dual Buffer)	99.95	42.47	10.78	—	—	475

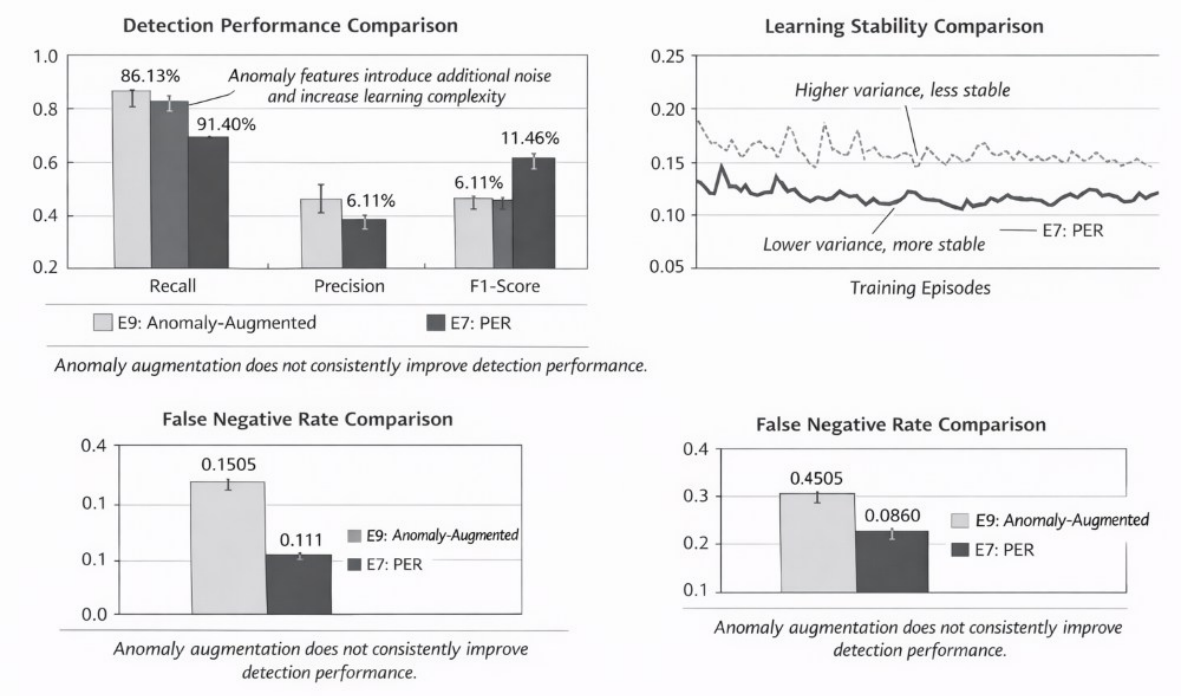


Figure 15. Performance Comparison: E7 (PER Baseline) vs. E9 (Anomaly-Augmented RL)

4.6. Zero-Day Evaluation using LOAO (E8)

The LOAO zero-day evaluation does not require training the model on any attack family and only tests it on unseen samples, allowing estimation of its generalization performance to new attack types.

These results show a very clear distinction between known and zero-day conditions. The model has more than 96% recall for web-based and SQL attack families, but only about 1.6% precision, indicating a high false-positive rate under the known scenario. For some of the attacks, recall under zero-day conditions is also high: Brute Force XSS has 100% recall, SQL Injection has 97.70%, and Brute Force Web has 79.38%. All the above results are obtained with low false-positive rates.

However, the Infiltration attack type is a difficult case with a zero-day recall of 0.01%, meaning it is not detected under unseen conditions by the model. This implies complexity, or more fancifully, that pattern of attacks involving complex (or multi-staged) means may perhaps be the hardest to generalize.

https://doi.org/10.31436/iiumej.v27i2.4240

Table 11. LOAO Zero-Day Detection Results (E8).

Attack	Scenario	Precision (%)	Recall (%)	F1 (%)	FPR	FNR
Brute Force Web	Known	1.61	96.85	3.16	0.7135	0.0315
Brute Force Web	Zero-Day	—	79.38	—	—	0.2062
Brute Force XSS	Known	1.58	97.37	3.11	0.7313	0.0263
Brute Force XSS	Zero-Day	—	100.00	—	—	0.0000
SQL Injection	Known	1.59	97.23	3.13	0.7265	0.0277
SQL Injection	Zero-Day	—	97.70	—	—	0.0230
Infiltration	Known	49.50	26.88	34.84	0.0000	0.7312
Infiltration	Zero-Day	—	0.01	—	—	0.9999

Table 12. False Negative Rate Comparison: Known vs. Zero-Day Conditions (E8).

Attack	FN Rate (Known)	FN Rate (Zero-Day)
Brute Force Web	0.0315	0.2062
Brute Force XSS	0.0263	0.0000
SQL Injection	0.0277	0.0230
Infiltration	0.7312	0.9999

Table 13. Window-Based Stability Metrics under LOAO Evaluation (E8).

Attack	Mean Recall	Variance	Std Dev
Brute Force Web	0.9687	5.72×10^{-6}	0.00239
Brute Force XSS	0.9735	6.28×10^{-6}	0.00251
SQL Injection	0.9723	4.94×10^{-6}	0.00222
Infiltration	0.2588	8.44×10^{-3}	0.09185

Web-based and SQL attack families, the recall variance is of the order of 10^{-6}. Therefore, in these families, we can conclude that the detection remains stable across different evaluation instances. The variance and standard deviation for Infiltration are 8.44×10^{-3} and 0.0918, implying decreased consistency in detection performance as compared to the first threat in this domain. In general, the performance results show that the model generalizes well to certain families of attacks but remains weak against more complex or stealthy attack patterns.

4.7. Drift Evaluation and Stability Analysis (E10–E11)

Experiments E10 and E11 evaluate robustness to distributional shift using a window-based evaluation framework applied to sequential traffic segments with varying attack proportions. E10 introduces synthetic drift through controlled window configurations, while E11 evaluates real drift by applying the E7 policy to an unseen temporal traffic partition.

4.7.1. Performance under Synthetic Drift (E10)

Under synthetic drift, the model maintains high attack recall across all evaluation windows (95.74–98.00%), with false negatives not exceeding two per window. However, false positive counts exceed 92,000 per window, resulting in ARMF values of approximately 137,000–

138,000 alerts per million flows. This indicates a detection behavior characterized by high recall accompanied by a substantial alert volume.

Table 14. Window-Wise Performance Metrics under Synthetic Drift (E10).

Window	TP	FP	FN	TN	Recall (%)	Prec (%)	ARMF	Mean Reward
Global	180	370,942	6	2,326,000	96.77	0.049	137,599	0.2938
Win 1	45	92,680	2	581,555	95.74	0.049	137,517	0.2939
Win 2	53	93,196	2	581,031	96.36	0.057	138,294	0.2928
Win 3	33	92,682	1	581,566	97.06	0.036	137,502	0.2939
Win 4	49	92,384	1	581,848	98.00	0.053	137,084	0.2946

4.7.2. Performance under Real Drift (E11)

The model's performance profile changes when real drift is present. Recall is stable across all evaluation windows, ranging from about 89% to 92% (variance=1.59×10−4, std dev=0.013). ARMF drops considerably compared to E10, from 1,011 to 1,471 alerts per million flows (mean=1,189), with the number of false positives per window reduced to about 640–682. The results show that under realistic traffic conditions, detection behavior becomes more balanced: recall remains stable, while alert volume decreases.

Table 15. Window-Wise Performance Metrics under Real Drift (E11).

Scope	Recall (%)	Prec (%)	F1 (%)	FP	FN	ARMF
Global Stream	91.27	18.58	30.88	2,611	57	1,189
Window 1	89.36	6.16	11.52	640	5	1,011
Window 2	91.99	31.25	46.65	682	27	1,471
Window 3	91.09	22.33	35.87	640	18	1,222
Window 4	89.55	8.46	15.46	649	7	1,051

4.7.3. Drift Robustness and Comparative Analysis

These two operational behaviors can be illustrated by comparing E10 and E11. In the case of synthetic drift with an extremely low attack ratio, your model yields higher recall but an overabundance of alerts. Conversely, the model achieves detection at a very low drift level, reduces alert volume, and performs well when agnostic to drift.

The low level of temporal variability in scores across evaluation windows results in an ARMF variance of 43,720 and a recall variance of 1.59×10^{-4} (E11). These findings imply that the learned policy is invariant to changes in knowledge distribution.

https://doi.org/10.31436/iiumej.v27i2.4240

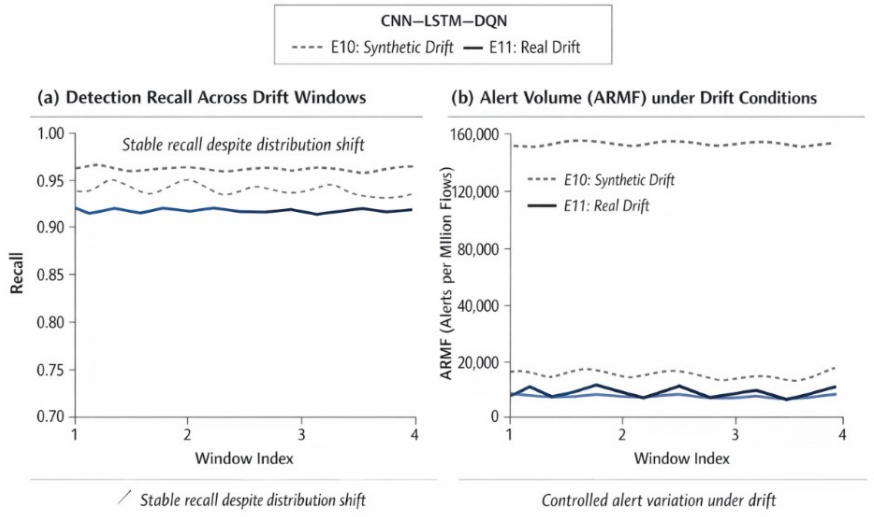


Figure 16. Drift Window Performance: recall and ARMF trajectories across E10 and E11 windows.

4.7.4. Implication for Real-World Deployment

The results for E11 show that our proposed framework, combined with ARMF, yields about 90% recall, at an average of~ 1000 - 1500 alerts per million of flows. This trade-off between detection performance and alert volume is relevant to deployment-focused intrusion detection, where excessive false positives can worsen the effectiveness of deployed sensors. The stability across the evaluation windows also indicates that, with such a framework, we can expect reasonable performance in scenarios where traffic distributions evolve over time.

4.8. Stability-Aware Evaluation using ARMF

ARMF (above) is employed as a critical operational performance indicator for wetland alert volume across all experimental phases. Conventional metrics based on accuracy may not accurately reflect models' performance in such an extreme class-imbalance setting (attack ratio < 0.01%). ARMF, on the other hand, offers an appealingly straightforward metric of alert burden that is critical in deployed intrusion detection environments.

Table 16. ARMF and Recall Summary across Experimental Stages.

Experiment	Model	Recall (%)	ARMF	FNR
E1	Baseline CNN	98.39	44,845	0.016
E2	Baseline LSTM	40.86	158,150	0.591
E3	CNN–LSTM	97.31	31,413	0.027
E4	CNN–LSTM Warm-Start	98.39	32,858	0.016
E5	DQN Naive RL	42.47	383	0.575
E6	DQN + Replay	84.95	12,728	0.151
E7	DQN + PER	91.40	1,031	0.086
E11	E7 Policy (Real Drift)	91.27	1,189	0.095

The trajectory from E1 to E7 demonstrates a common pattern: supervised models have high recall but generate significant alerts; naive reinforcement learning sacrifices alert volume for recall; and PER-augmented reinforcement learning (E7) achieves a better balance between detection effectiveness and alert volume. The detection threshold analysis provides additional information: the system's alert volume and sensitivity can be tuned by threshold selection, enabling the operator to select operating points for various deployed scenarios and risk tolerances.

https://doi.org/10.31436/iiumej.v27i2.4240

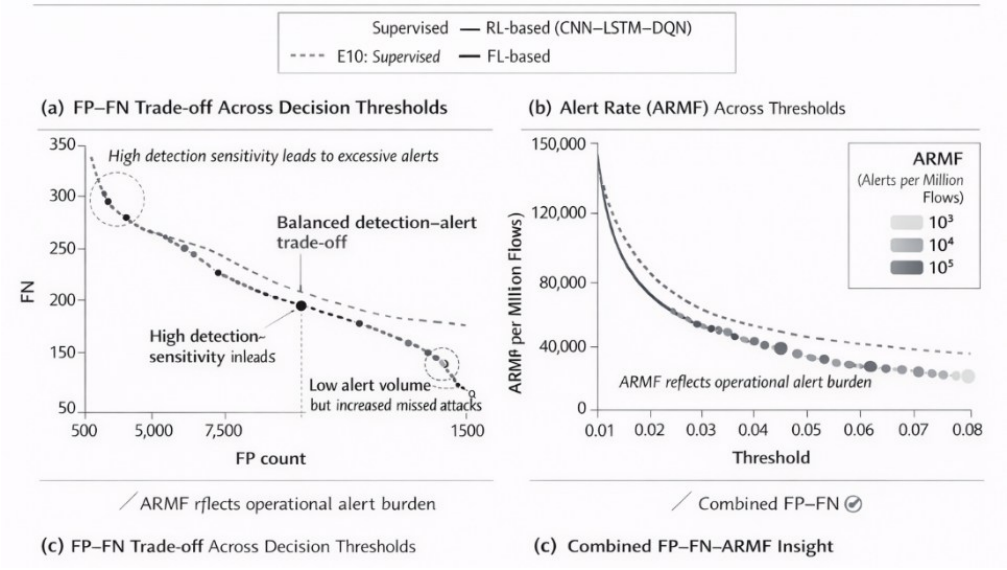


Figure 17. FP–FN Trade-off Curve under Stability-Aware Evaluation (ARMF) for E10

5. Threats to Validity and Discussion of Limitations

5.1. Internal Validity

The main internal validity issue concerns the distribution difference between the reinforcement learning training set and the entire evaluation dataset. The RL training stages (E4-E9) utilize a balanced subset of 500,742 flows designed for stable policy learning under severe class imbalance, and each evaluation stage is performed on the full imbalanced test set. The difference between the training and development distributions may inhibit generalization when using the learned policy with traffic distributions different from those in the training subset.

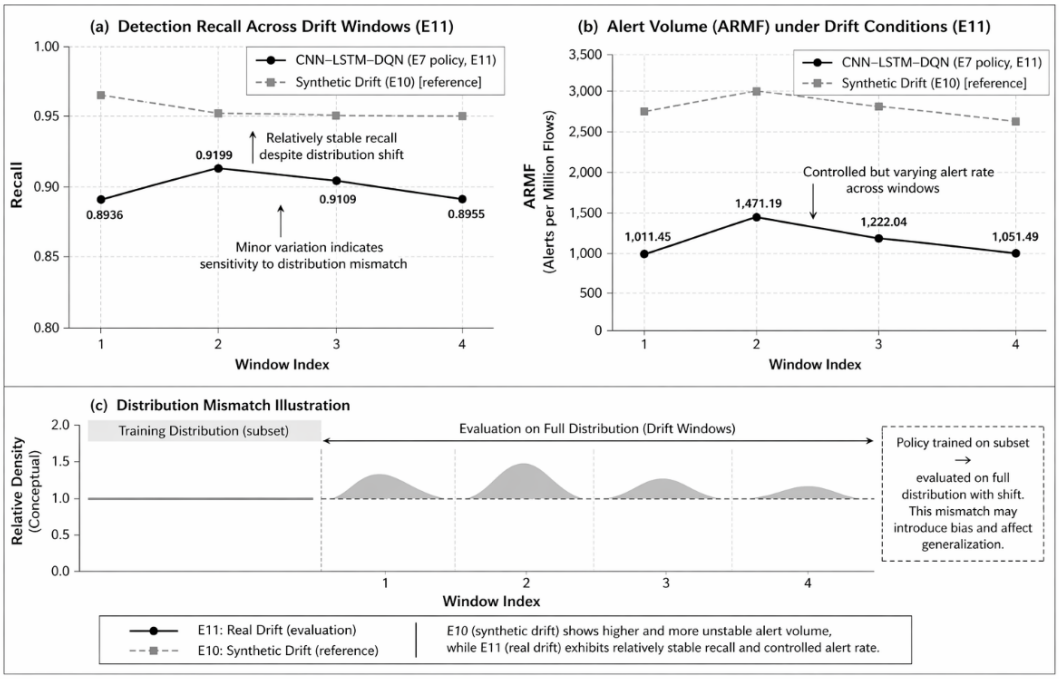


Figure 18. Drift Window Performance under Distribution Mismatch (E10-E11)

The influence is explained in Section 4.7 by the findings of the examination of the alignments across the datasets (drift evaluation). For synthetic drift (E10), the model exhibits high recall and an ARMF of ~137k. Conversely, in true drift settings (E11), ARMF (mean: 1,189) and recall are much enhanced, with ARMF being essentially unchanged. This suggests that the learned policy varies across different distributional settings. In addition, the reported variance measures capture the variability of recall and alert rates across assessment windows, indicating that the training-evaluation distribution shift creates quantifiable uncertainty. You must take this influence into account when interpreting the stability results.

5.2. Computational Complexity and Practical Constraints

The computational overhead of the CNN–LSTM–DQN–PER framework is greater than that of traditional supervised IDS algorithms. This architecture consists of 1D convolutional layers and bidirectional LSTM modules, with reinforcement learning components (experience replay buffers, target networks, and prioritized sampling mechanisms). The RL stages (E5–E7) take about 200–1,290 seconds to train, whereas the supervised baselines (E1–E3) require 66–99 seconds, with corresponding brain population sizes.

Although inference time is shorter than training time, because training can still take days to weeks or longer, this means that full 3D segmentation is not applicable in resource-constrained environments. You might also investigate model compression and pruning, or hardware acceleration, to reduce computational load.

5.3. External Validity

The evaluation is based only on the CSE-CIC-IDS2018 dataset, which is also simulated rather than a realistic operational network, even though it is widely used and diverse in attack classes. Different organizations and infrastructures may have different traffic patterns, attack distributions, and defense or deployment approaches. Consequently, the proposed framework may not be generalizable to other datasets or deployment contexts. Cross-dataset validation and evaluation on real-world network traffic are important open directions for future work.

5.4. Limitations of Reinforcement Learning-Based IDS

Using deep reinforcement learning to build IDS introduces specific design considerations. DQN performance is sensitive to the design of the reward function, as evident in Section 4.4. The naive RL setting (E5) yields the most conservative detection behavior by reducing the predicted attack frequency, resulting in a false negative rate of 57.5% while maintaining high accuracy under imbalanced conditions.

The results from the anomaly-augmented experiments presented in Section 4.5 also suggest that the scale of representation does not always correlate with a higher detection performance. Moving to E9 and E9b, the autoencoder anomaly scores incorporated led to a recall loss of 91.40% (E7) −> 42.47%. These results indicate that boosting the state representation may make learning more difficult, negatively impacting performance. These results suggest that careful design of the reward function and selection of state representations may be important for the success of reinforcement learning–based IDS.

https://doi.org/10.31436/iiumej.v27i2.4240

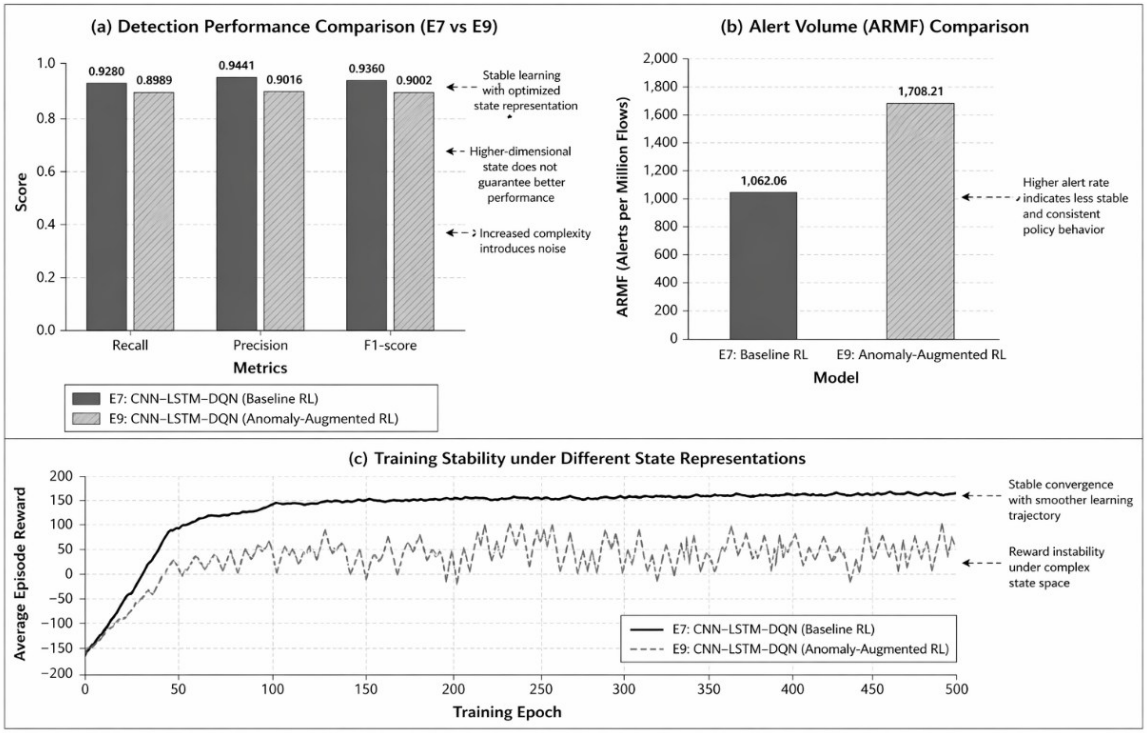


Figure 19. Comparison of E7 (PER Baseline) vs. E9 (Anomaly-Augmented RL): recall, ARMF, and training reward convergence

5.5. Limitations in Evaluation Scope

Several evaluation limitations are noted. First, the study does not include direct benchmarking against external state-of-the-art IDS models under the same dataset and evaluation protocol, which would strengthen the comparative assessment of the proposed framework. Second, stability-aware metrics are primarily reported for reinforcement learning–based models in this study; extending these metrics consistently across all experimental stages, including supervised baselines, would provide a more comprehensive comparison. Third, the drift evaluation is based on windowed partitions rather than live traffic streams, which may limit the extent to which the results reflect real-world deployment conditions.

5.6. Summary of Limitations

In summary, the limitations discussed in this study highlight the general difficulties in designing and evaluating intrusion detection systems in a dynamic network environment. Although this framework successfully tackles several important challenges, such as adaptivity, class imbalance, and stability, generalization (across broader deployment scenarios), computational efficiency, and evaluation opportunities exist primarily because of limitations in current simulator design, which allow us to capture only a narrow window of realistic deployments.

6. Conclusions and Future Work

6.1 Conclusion

This study proposed a hybrid CNN–LSTM–DQN–PER intrusion detection framework together with a stability-aware evaluation methodology to address the limitations of conventional IDS under zero-day attacks, severe class imbalance, and non-stationary traffic

conditions. The framework integrates spatio-temporal feature learning with reinforcement learning–based adaptive policy optimization, enabling a dynamic trade-off between detection performance and alert volume.

Experimental results consistently demonstrate that conventional supervised models achieve high recall (up to 98.39%) but produce excessive alert volumes, with ARMF reaching up to 44,845 under imbalanced conditions. In contrast, the proposed reinforcement learning with prioritized experience replay (E7) shows more balanced performance, with a recall of 91.40% and an ARMF of 1,031. Compared to naive reinforcement learning (E5), which yields a recall of 42.47%, the proposed PER-based approach significantly improves detection capability while preserving low alert rates.

Across reinforcement learning stages (E5–E7), stabilization mechanisms improve both detection performance and learning consistency. The model maintains stable recall (around 89–92%) under real-world drift conditions (E11) with a controlled alert volume (ARMF ≈ 1,189), demonstrating robustness to varying traffic distributions. Zero-day evaluation with the LOAO protocol further suggests that the model generalizes to other types of unseen attacks, with high recall across different attack categories, while performance remains limited in more complex scenarios such as infiltration. These results shed light on the merits and limitations of the proposed method.

In summary, the results demonstrate that there is a trade-off among detection accuracy, alert volume, and temporal stability in effective IDS design, and the proposed stability-aware evaluation framework, based on ARMF and FP–FN trade-off analysis, provides a more deployment-oriented assessment than conventional accuracy-based metrics.

6.2. Limitations Reflection

This study offers several contributions, but it has a few limitations (see Section 5). Initially, this approach depends on a single dataset, which may restrict the generalizability of the findings. Moreover, the reinforcement learning training setup uses a balanced subset, which leads to a distribution mismatch between training and evaluation conditions. Second, while stability-aware metrics were incorporated into the evaluation framework, they are not uniformly applied across experimental configurations. Such restrictions on direct comparison with external state-of-the-art IDS models, in addition to potential benchmarking against existing approaches.

Third, the computational cost of the proposed framework could limit its practical implementation in resource-constrained systems. Admittedly, these limitations indicate that, although the framework provides a means of investigation for adaptive IDS design, further work on generalization, computational cost, and comparative assessment will be required.

6.3. Future Work

There are various limitations of the approach that will be addressed in future work, along with possible extensions to our framework. Firstly, a cross-dataset validation is conducted to assess model generalization across different network environments and traffic distributions using the CSE-CIC-IDS2018 dataset. Secondly, it will examine lightweight model architectures and optimization techniques to reduce computational overhead and improve efficiency. These include model compression, pruning, and hardware acceleration for efficient deployment in constrained environments. In addition, we will look more deeply into reward design and reinforcement learning strategies that lead to training stability and sample efficiency. Specific focus will be given to adaptive reward formulations that maximize detection performance in a changing environment, while reducing alert volume.

Finally, further work will extend the evaluation framework to facilitate comparisons with state-of-the-art IDS models based on studies external to this dissertation and provide a means for deployment-oriented evaluation in real-world network environments. This will give you a better evaluation of how the framework performs under practical cybersecurity conditions.

REFERENCES

[1] A. Halbouni, T. S. Gunawan, M. H. Habaebi, M. Halbouni, M. Kartiwi, and R. Ahmad, "CNN-LSTM: Hybrid Deep Neural Network for Network Intrusion Detection System," *Ieee Access*, vol. 10, pp. 99837–99849, 2022, doi: 10.1109/access.2022.3206425.

[2] M. Abdallah, N. Le-Khac, H. Z. Jahromi, and A. D. Jurcut, "A Hybrid CNN-LSTM Based Approach for Anomaly Detection Systems in SDNs," pp. 1–7, 2021, doi: 10.1145/3465481.3469190.

[3] H. C. Altunay and Z. Albayrak, "A hybrid CNN+LSTM-based intrusion detection system for industrial IoT networks," *Eng. Sci. Technol. an Int. J.*, vol. 38, p. 101322, 2023, doi: https://doi.org/10.1016/j.jestch.2022.101322.

[4] Y. D. Lin, H. X. Huang, D. Sudyana, and Y. C. Lai, "AI for AI-based intrusion detection as a service: Reinforcement learning to configure models, tasks, and capacities," *J. Netw. Comput. Appl.*, vol. 229, no. February, p. 103936, 2024, doi: 10.1016/j.jnca.2024.103936.

[5] M. Alrehaili and A. Alshamrani, "A Hybrid Deep Learning Approach for Advanced Persistent Threat Attack Detection," pp. 78–86.

[6] C. Do Xuan and M. H. Dao, "A novel approach for APT attack detection based on combined deep learning model," *Neural Comput. Appl.*, vol. 33, no. 20, pp. 13251–13264, 2021, doi: 10.1007/s00521-021-05952-5.

[7] [N. K. Almazmomi, "Advanced Persistent Threat Detection Using Optimized and Hybrid Deep Learning Approach," *Secur. Priv.*, vol. 8, no. 2, p. e70011, Mar. 2025, doi: https://doi.org/10.1002/spy2.70011.

[8] K. Alam, M. F. Monir, M. J. Hossain, M. S. Uddin, and M. T. Habib, "Adaptive Defense: Zero-Day Attack Detection in NIDS With Deep Reinforcement Learning," *IEEE Access*, vol. 13, pp. 116345–116361, 2025, doi: 10.1109/ACCESS.2025.3585445.

[9] E. H. Omoush, M. Almseidin, and A. Aldweesh, "A Self-Adaptive Intrusion Detection System for Zero-Day Attacks Using Deep Q-Networks," *IEEE Access*, 2025.

[10] V. Sharma, "Rainbow dqn for intrusion detection: A unified deep reinforcement learning approach across benchmark datasets," *Int. J. Appl. Math.*, vol. 38, no. 5s, pp. 647–675, 2025.

[11] M. Alkasassbeh, E. H. Omoush, M. Almseidin, and A. Aldweesh, "A Self-Adaptive Intrusion Detection System for Zero-Day Attacks Using Deep Q-Networks," *IEEE Access*, vol. 13, no. August, pp. 174280–174296, 2025, doi: 10.1109/ACCESS.2025.3617792.

[12] Y. Wu, Y. Hu, J. Wang, M. Feng, A. Dong, and Y. Yang, "An active learning framework using deep Q-network for zero-day attack detection," *Comput. Secur.*, vol. 139, p. 103713, Apr. 2024, doi: 10.1016/j.cose.2024.103713.

[13] H. A. Sakr, M. M. Fouda, A. F. Ashour, A. Abdelhafeez, M. I. El-Afifi, and M. Refaat Abdellah, "Machine learning-based detection of DDoS attacks on IoT devices in multi-energy systems," *Egypt. Informatics J.*, vol. 28, no. May, p. 100540, 2024, doi: 10.1016/j.eij.2024.100540.

[14] I. H. Sarker, CyberLearning: Effectiveness analysis of machine learning security modeling to detect cyber-anomalies and multi-attacks, vol. 14, no. April. 2021. doi: 10.1016/j.iot.2021.100393.

[15] W. Ren *et al.*, "APT Attack Detection Based on Graph Convolutional Neural Networks," *Int. J. Comput. Intell. Syst.*, vol. 16, no. 1, p. 184, 2023, doi: 10.1007/s44196-023-00369-5.

[16] L. Y. Por *et al.*, "A Systematic Literature Review on the Methods and Challenges in Detecting Zero-Day Attacks: Insights from the Recent CrowdStrike Incident," *IEEE Access*, vol. 12, no. October, pp. 144150–144163, 2024, doi: 10.1109/ACCESS.2024.3455410.

[17] R. Article, M. M. Issa, M. Aljanabi, and H. M. Muhialdeen, "Systematic literature review on intrusion detection systems : Research trends , algorithms , methods , datasets , and limitations," 2024.

[18] Rushendra, K. Ramli, N. Hayati, E. Ihsanto, T. S. Gunawan, and A. H. Halbouni, "Development of Intrusion Detection System using Residual Feedforward Neural Network Algorithm," *2021 4th Int. Semin. Res. Inf. Technol. Intell. Syst. ISRITI 2021*, pp. 539–543, 2021, doi: 10.1109/ISRITI54043.2021.9702773.

[19] J. A. Shaikh *et al.*, "RCLNet: an effective anomaly-based intrusion detection for securing the IoMT system," *Front. Digit. Heal.*, vol. 6, no. October, pp. 1–12, 2024, doi: 10.3389/fdgth.2024.1467241.

[20] J. A. Shaikh *et al.*, "A deep Reinforcement learning-based robust Intrusion Detection System for securing IoMT Healthcare Networks," *Front. Med.*, vol. 12, no. April, 2025, doi: 10.3389/fmed.2025.1524286.

[21] I. Technology *et al.*, "A Systematic Literature Review of Intrusion Detection System for Network Security : Research Trends , Datasets and Methods," no. May, pp. 0–5, 2020, doi: 10.1109/ICICoS51170.2020.9299068.

[22] S. S. Bamber, A. V. R. Katkuri, S. Sharma, and M. Angurala, "A hybrid CNN-LSTM approach for intelligent cyber intrusion detection system," *Comput. Secur.*, vol. 148, no. June 2024, p. 104146, 2025, doi: 10.1016/j.cose.2024.104146.

[23] S. Elsayed, K. Mohamed, and M. A. Madkour, "A Comparative Study of Using Deep Learning Algorithms in Network Intrusion Detection," *IEEE Access*, vol. 12, no. May, pp. 58851–58870, 2024, doi: 10.1109/ACCESS.2024.3389096.

[24] P. Sinha, D. Sahu, S. Prakash, T. Yang, R. S. Rathore, and V. K. Pandey, "A high performance hybrid LSTM CNN secure architecture for IoT environments using deep learning," *Sci. Rep.*, vol. 15, no. 1, pp. 1–26, 2025, doi: 10.1038/s41598-025-94500-5.

[25] U. Agarwal, "Reinforcement Learning and Hybrid CNN-LSTM Based Host-Intrusion Detection System," no. April, 2025, doi: 10.13140/RG.2.2.24939.25129.

[26] A. H. Halbouni, T. S. Gunawan, M. Halbouni, F. A. A. Assaig, M. R. Effendi, and N. Ismail, "CNN-IDS: Convolutional Neural Network for Network Intrusion Detection System," in *Proceeding of 2022 8th International Conference on Wireless and Telematics, ICWT 2022*, 2022, pp. 1–4. doi: 10.1109/ICWT55831.2022.9935478.

[27] C. Do Xuan and T. T. Nguyen, "A novel approach for APT attack detection based on an advanced computing," pp. 1–19, 2024.

[28] N. Saini, V. Bhat Kasaragod, K. Prakasha, and A. K. Das, "A hybrid ensemble machine learning model for detecting APT attacks based on network behavior anomaly detection," *Concurr. Comput. Pract. Exp.*, vol. 35, no. 28, pp. 1–27, 2023, doi: 10.1002/cpe.7865.

[29] Youaccell, "Reinforcement Learning: Adaptive Security Measures," Youaccell. [Online]. Available: https://youaccel.com/lesson/reinforcement-learning-adaptive-security-measures/premium

[30] M. R. Naeem, R. Amin, M. Farhan, F. S. Alsubaei, E. Alsolami, and M. D. Zakaria, "Cyber security Enhancements with reinforcement learning: A zero-day vulnerabilityu identification perspective," *PLoS One*, vol. 20, no. 5, p. e0324595, May 2025, [Online]. Available: https://doi.org/10.1371/journal.pone.0324595

[31] M. R. Naeem, R. Amin, M. Farhan, F. S. Alsubaei, E. Alsolami, and M. D. Zakaria, "Cyber security Enhancements with reinforcement learning: A zero-day vulnerabilityu identification perspective.," *PLoS One*, vol. 20, no. 5, p. e0324595, 2025, doi: 10.1371/journal.pone.0324595.

[32] Y. Wu, Y. Hu, J. Wang, M. Feng, A. Dong, and Y. Yang, "An active learning framework using deep Q-network for zero-day attack detection," *Comput. Secur.*, vol. 139, no. December 2023, p. 103713, 2024, doi: 10.1016/j.cose.2024.103713.

[33] F. Ullah, S. Ullah, G. Srivastava, and J. C.-W. Lin, "IDS-INT: Intrusion detection system using transformer-based transfer learning for imbalanced network traffic," *Digit. Commun. Networks*, vol. 10, no. 1, pp. 190–204, 2024, doi: https://doi.org/10.1016/j.dcan.2023.03.008.

[34] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization.," *ICISSp*, vol. 1, no. 2018, pp. 108–116, 2018.

[35] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in *2010 IEEE symposium on security and privacy*, IEEE, 2010, pp. 305–316.

[36] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.

[37] J. G. Fiscus and G. R. Doddington, "Topic detection and tracking evaluation overview," in *Topic detection and tracking: event-based information organization*, Springer, 2002, pp. 17–31.

[38] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," *ACM Comput. Surv.*, vol. 46, no. 4, pp. 1–37, 2014.

[39] A. Bifet and R. Gavalda, "Learning from time-changing data with adaptive windowing," in *Proceedings of the 2007 SIAM international conference on data mining*, SIAM, 2007, pp. 443–448.

[40] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*, vol. 1, no. 1. MIT press Cambridge, 1998.

https://doi.org/10.31436/iiumej.v27i2.4246

A Robust Low-Complexity Star Centroiding Algorithm for Autonomous Navigation under Lunar Noise Conditions

ANIS HANNANI RAZAMAN[1], YASSER ASRUL AHMAD[1*], TEDDY SURYA GUNAWAN[1], OTHMAN OMRAN KHALIFA[1], MILA POPOVIĆ[2]

[1]*Department of Electrical and Computer Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia*
[2]*Astronomical Observatory Belgrade, Serbia*

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

(Received: 27 February 2026; Accepted: 1 May 2026; Published online: 10 May 2026)

ABSTRACT: Star-based navigation on the lunar surface is severely degraded by strong regolith reflections, abrupt illumination transitions, and sensor-induced noise, all of which lower the signal-to-noise ratio (SNR) and impair centroiding accuracy. Existing methods address this trade-off poorly: classical center-of-mass (COM) and Gaussian fitting are computationally light but noise-sensitive, whereas iterative weighting and learning-based approaches improve accuracy at the cost of high computational load and large training data. This study addresses this gap by proposing a low-complexity yet robust star centroiding algorithm tailored for lunar-surface imagery. The pipeline integrates median filtering for impulse-noise suppression, an adaptive global threshold (set at 30% of peak intensity) for star-region segmentation, and an intensity-weighted COM estimator for sub-pixel localization. The method was implemented in MATLAB R2023b and evaluated on 30 Stellarium-derived star fields, each corrupted by Gaussian, Poisson, salt-and-pepper, speckle, and solar-glare noise spanning SNRs from −4.71 dB to 0.72 dB. Benchmarked against standard COM, Gaussian fitting, and the Sieve Search Algorithm (SSA), the proposed method achieves the lowest average root-mean-square error (RMSE = 1.218 pixels), the lowest Euclidean distance error (1.143 pixels), and the lowest false detection rate (FDR = 6.716%), corresponding to relative reductions of 21.5%, 16.7%, and 32.7% over the best baseline, respectively. The algorithm's favorable accuracy–complexity trade-off makes it well-suited for resource-constrained onboard processors in future lunar exploration and autonomous spacecraft navigation missions.

ABSTRAK: Navigasi berasaskan bintang di permukaan bulan terjejas dengan ketara akibat pantulan regolit yang kuat, perubahan pencahayaan yang mendadak, dan hingar daripada penderia, yang kesemuanya menurunkan nisbah isyarat kepada hingar (SNR) dan menjejaskan ketepatan penganggaran sentroid. Kaedah sedia ada gagal mengimbangi keseimbangan ini: kaedah pusat jisim (COM) klasik dan pemadanan Gaussian adalah ringan dari segi pengiraan tetapi sensitif terhadap hingar, manakala kaedah pemberat berulang dan berasaskan pembelajaran mesin meningkatkan ketepatan tetapi memerlukan beban pengiraan tinggi dan data latihan yang besar. Kajian ini mengisi jurang tersebut dengan mencadangkan satu algoritma penganggaran sentroid bintang yang teguh dan berkompleksiti rendah, khusus untuk imej permukaan bulan. Saluran pemprosesan menggabungkan penapisan median untuk menindas hingar impuls, ambang global mudah suai (ditetapkan pada 30% daripada keamatan puncak) untuk segmentasi kawasan bintang, dan penganggar COM berwajaran keamatan untuk penyetempatan subpiksel. Algoritma dilaksanakan dalam MATLAB R2023b dan dinilai ke atas 30 medan bintang Stellarium yang dicemari oleh hingar Gaussian, Poisson, garam-dan-lada, bintik, dan silau solar dengan SNR dari −4.71 dB hingga 0.72 dB. Berbanding COM standard, pemadanan Gaussian, dan Algoritma Carian Ayak (SSA), kaedah

https://doi.org/10.31436/iiumej.v27i2.4246

yang dicadangkan mencapai purata RMSE terendah (1.218 piksel), jarak Euclidean terendah (1.143 piksel), dan kadar pengesanan palsu terendah (FDR = 6.716%), masing-masing mewakili pengurangan relatif 21.5%, 16.7%, dan 32.7% berbanding garis dasar terbaik. Imbangan ketepatan-kompleksiti yang baik menjadikan algoritma ini sesuai untuk pemproses dalam-pesawat dengan sumber terhad bagi misi penerokaan bulan dan navigasi pesawat angkasa autonomi pada masa hadapan.

KEYWORDS: *star centroiding, lunar navigation, harsh lunar environment, image processing, autonomous navigation.*

1. INTRODUCTION

Star sensors are essential components of autonomous spacecraft navigation. By identifying stellar patterns in captured images, they enable attitude determination with arc-second precision, providing a passive, drift-free reference that is independent of GPS, ground links, or planetary line-of-sight. Compared with celestial bodies such as Earth, the Sun, or the Moon, the abundance and angular stability of cataloged stars make them an ideal reference for orientation calculations [1]. With renewed international interest in lunar exploration, including missions under the International Lunar Research Station (ILRS) and the Turkish Lunar Mission [2], [3], [4], star-based navigation has re-emerged as a promising solution for surface and orbital operations. Historically, star trackers supported early lunar missions: the Apollo Command and Lunar Modules relied on optical star sightings for trajectory corrections and lunar descent navigation [5]. More recently, China's Chang'e-3 and Chang'e-4 landers used star trackers during transfer-orbit and lunar-orbit insertion phases, and the Queqiao relay satellite employed them to maintain Earth–Moon L2 halo orbit operations [6], [7].

Despite this heritage, the lunar surface remains an exceptionally hostile imaging environment for star sensors. The absence of an atmosphere, combined with strong solar reflection from the regolith, electrostatic dust levitation, and abrupt illumination transitions across the terminator, drives the SNR of star imagery into the negative-decibel regime, where centroiding errors and false detections rise sharply. Existing solutions lie at the two extremes of an unresolved trade-off. On one hand, classical centroiding methods such as the standard Center of Mass (COM), Gaussian fitting, and the recently proposed Sieve Search Algorithm (SSA) are computationally light and easy to implement, but degrade rapidly under low-SNR lunar conditions. On the other hand, learning-based and iterative-weighting methods improve robustness but at the cost of substantial computational load, parameter tuning, or large training datasets [8], [9], which are unattractive for the resource-constrained processors typically deployed on lunar landers and rovers.

To date, no published study has systematically benchmarked a low-complexity centroiding pipeline against multiple state-of-the-art baselines under a unified set of lunar-relevant noise models, including solar glare. This paper addresses that gap. The main contributions are threefold:

(i) a robust yet low-complexity centroiding pipeline integrating median filtering, an adaptive global threshold, and an intensity-weighted COM estimator, specifically tuned for lunar-surface imaging conditions;
(ii) a reproducible MATLAB-based evaluation framework using 30 Stellarium-derived star fields corrupted by five physically-motivated noise models (Gaussian, Poisson, salt-and-pepper, speckle, and solar glare) at SNRs ranging from −4.71 dB to 0.72 dB; and

https://doi.org/10.31436/iiumej.v27i2.4246

(iii) a quantitative comparison with COM, Gaussian fitting, and SSA, demonstrating consistent reductions in RMSE, Euclidean distance error, and false detection rate.

The remainder of this paper is organized as follows. Section 2 reviews the lunar imaging environment and prior centroiding methods. Section 3 details the proposed methodology and its MATLAB implementation. Section 4 presents the experimental results and critical analysis, and Section 5 concludes with directions for future work.

2. LITERATURE REVIEW

This section reviews the lunar imaging environment, the noise models commonly used to simulate it, and prior work on star detection and centroiding, identifying the limitations that motivate the proposed approach.

2.1. Moon Surface Environment

The lunar surface environment presents several physical and optical challenges that can affect the performance of star detection and centroiding for lunar-based navigation systems. The surface is mainly composed of two major geological regions: the highlands and the maria [10]. These variations in surface composition and reflectivity influence the background illumination captured by a star camera. The top layer, known as lunar regolith, is composed of fine dust, rock fragments, and glass particles produced by continuous micrometeoroid bombardment [11]. This regolith produces strong variations in reflection that contribute to background brightness and can interfere with star sensor imaging performance. The lunar regolith is present in the crust of the lunar layers, as shown in Figure 1, and the detailed sample of granular material that resembles dark, fine-grained sand or dirt is shown in Figure 2 [12], [13].

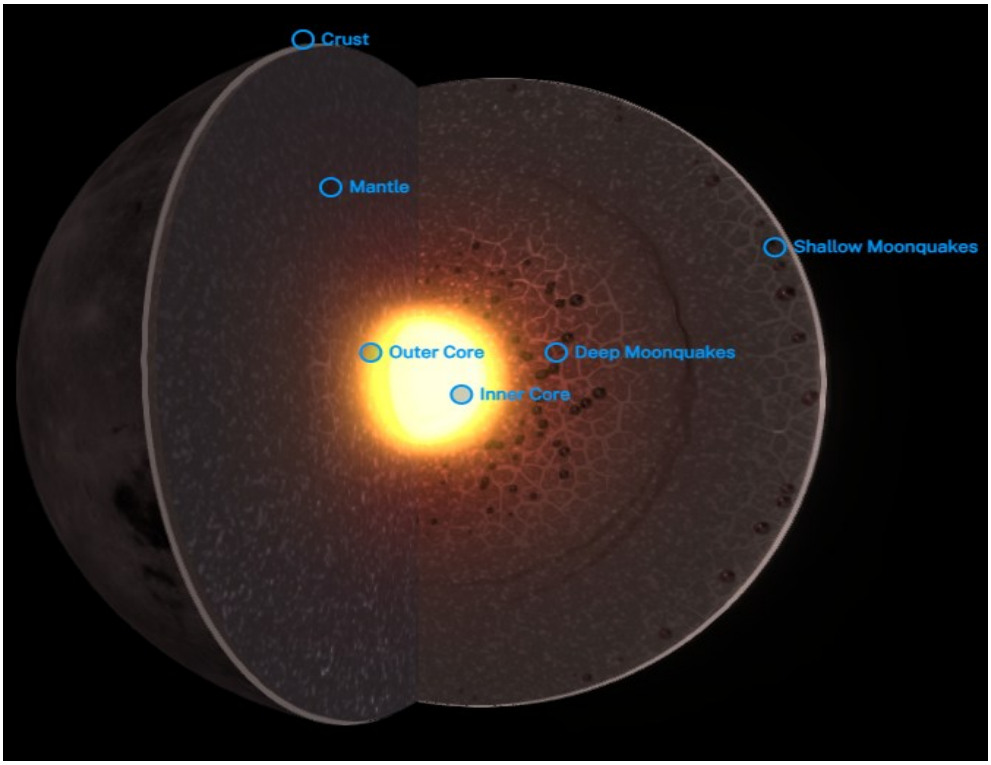


Figure 1. The regolith considered to be the uppermost layer of the lunar crust [13]

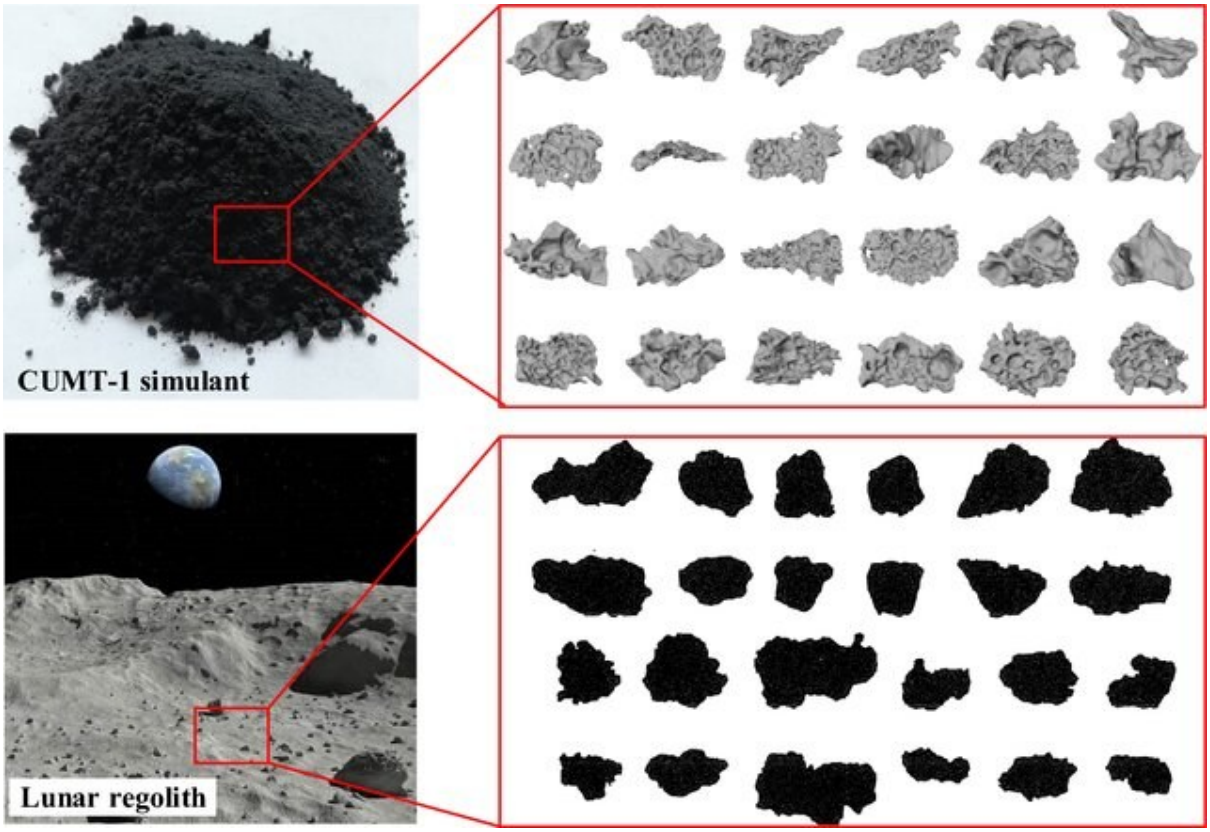


Figure 2. Detailed visualization of a lunar regolith simulant of CUMT-1 (China University of Mining and Technology-1) [12]

A defining characteristic of the lunar environment is the absence of a substantial atmosphere, so the sky remains completely black even during the day, with no atmospheric scattering. However, this condition also introduces several noise challenges. Sunlight reflecting off the regolith can generate glare, veiling luminance, and bright background gradients, all of which reduce the effective SNR for faint stars [14]. Additionally, electrostatic charging of lunar dust can lift particles above the surface, creating horizon glow and forward-scattering effects that introduce unwanted background illumination and intensity fluctuations in optical imaging systems [15], [16].

2.2. Different Types of Noise

The conditions outlined above translate into specific noise contributions that degrade star imagery. Faint stellar signals competing with strong, non-uniform regolith illumination, together with thermal and lighting extremes documented by NASA's Diviner instrument and the Apollo missions [17], [18], drive the captured SNR low, which directly reduces star visibility, increases false detections, and degrades centroid positional accuracy [19].

To recreate these lunar-surface challenges during algorithm development, several noise models are used to simulate degraded star fields. Table 1 describes the main noise types considered in this work, including Gaussian, Poisson, salt-and-pepper, and speckle noise [20], [21], [22], [23], as these represent common sensor- and photon-related disturbances in space imaging systems. Noise in digital images can generally be described as undesired brightness or color information, often visible as grainy texture or random intensity fluctuations [24]. Each

noise type arises from different physical or electronic mechanisms, and it is important to understand their properties for selecting appropriate noise reduction strategies.

Table 1. Noise and representations in the Lunar environment

Simulated Noises	Representations
Gaussian noise	Represents sensor electronic noise, such as thermal fluctuations in readout circuits, electronic sensor noise, and photon measurement randomness at low light levels.
Poisson noise	Simulates photon shot noise caused by the random nature of photon arrivals, which is amplified by the Moon's lack of atmosphere and the extreme contrast between illuminated and shadowed regions
Salt and pepper noise	Replicates impulse noise or high-contrast disruptions, specifically representing hardware/software failures, camera sensor defects, faulty pixels, or cosmic ray strikes
Speckle noise	Simulates multiplicative disturbances where noise scales with intensity, representing sensor gain fluctuations, optical imperfections, or rapid brightness changes caused by reflections from lunar dust or spacecraft surfaces
Light glare	Models intense solar reflections from the lunar regolith and direct solar illumination, which produce bright background gradients, veiling luminance, and structured light artifacts

2.3. Related Works on Star Detection

Two fundamental techniques used in image-based star detection are segmentation and centroiding. The segmentation technique removes background noise to isolate star information, and the centroiding technique is used to determine the precise positions of stars within an image [25], [26], [27]. These extracted positions are then utilized to estimate the orientation (attitude) of a spacecraft or satellite to enable precise navigation [28], [29].

Thresholding is the simplest and most widely used technique. It represents the first generation of low-level segmentation methods that rely on minimal prior information. Pixels with intensities above this threshold are classified as part of a star, while those below are treated as background noise [30], [31], [32]. According to Khalifa (2018), image segmentation involves grouping image regions that share similar characteristics to isolate relevant objects [24]. A recent study in medical imaging by Raman et al. introduced a mixed-thresholding and edge-preservation method to denoise MRI brain images while maintaining the important structural details in the medical image [33]. The thresholding dynamically responds to the noise characteristics, and next enables strong noise reduction without disturbing important image features [32], [34].

After the detection, the centroiding mechanism is applied to determine the precise location of each star. The process involves computing the intensity-weighted average of pixel positions within a star region, and finally results in a sub-pixel estimation of the star's center [35], [36], [37]. According to Liebe, the accuracy of a star tracker is strongly influenced by centroiding precision, particularly in the presence of image noise, pixel quantization effects, and optical distortions [19]. A novel technique, the Sieve Search Centroiding Algorithm (SSA) [38], improves accuracy by dividing images into smaller sub-regions and analyzing intensity distributions to compute the Center of Mass (COM). While it offers speed advantages over complex fitting algorithms, its sensitivity to noise and parameter variations remains a challenge. Figure 3 shows the graph of centroiding accuracy and the error curve from this study.

https://doi.org/10.31436/iiumej.v27i2.4246

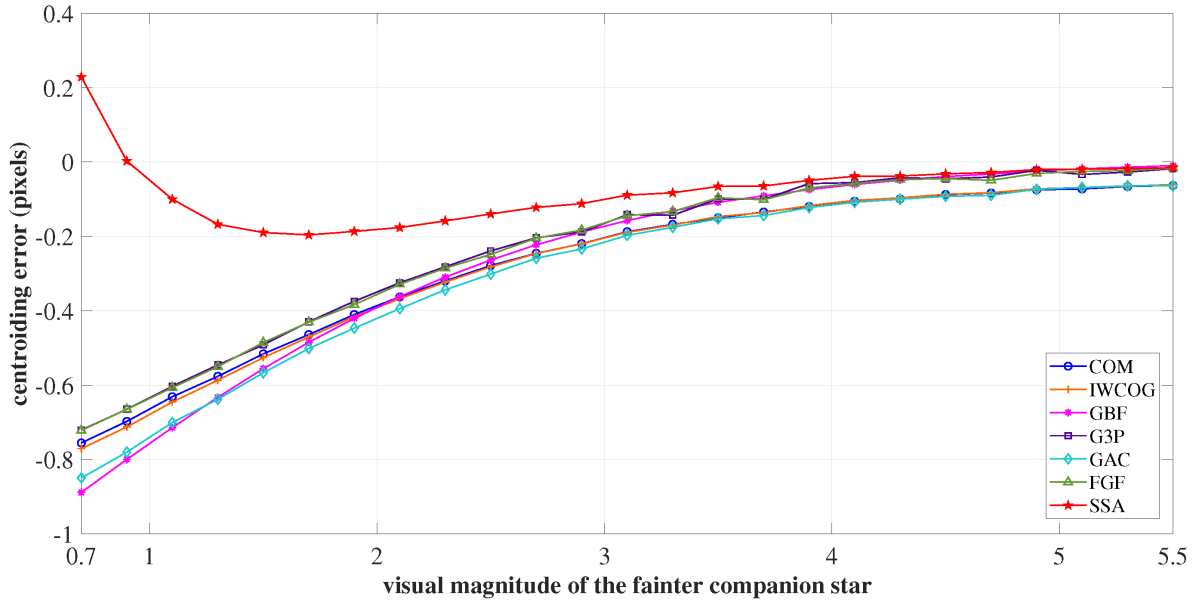


Figure 3. Centroiding error in estimating the brightness barycentre of an optical double-star pair [38]

Figure 4 illustrates how centroiding accuracy varies as the brightness barycentre shifts between the primary and secondary stars. When the secondary star is fainter than 4 VM (visual magnitude), most methods, including COM, Gaussian best fit (GBF), Gaussian three-point centroiding (G3P), Fast Gaussian Fitting (FGF), and SSA, estimate the centroid accurately, while iterative weighted center-of-gravity (IWCOG) and Gaussian Analytic Centroiding (GAC) show noticeable errors due to sensor noise and photo-electron spillover. While SSA is much faster than fitting methods, it is still inherently slower than the simplest grey-scale algorithms, such as COM or GAC.

A study introduced a dual-stage algorithm combining object segmentation and centroid extraction to eliminate non-stellar bright objects [26]. The primary contribution of this research is an effective pre-processing workflow designed to mitigate the challenge posed by non-stellar bright objects appearing in the star tracker's field of view. The centroiding technique proposed relies either on the Center of Gravity (CoG) or on locating the pixel of maximum intensity.



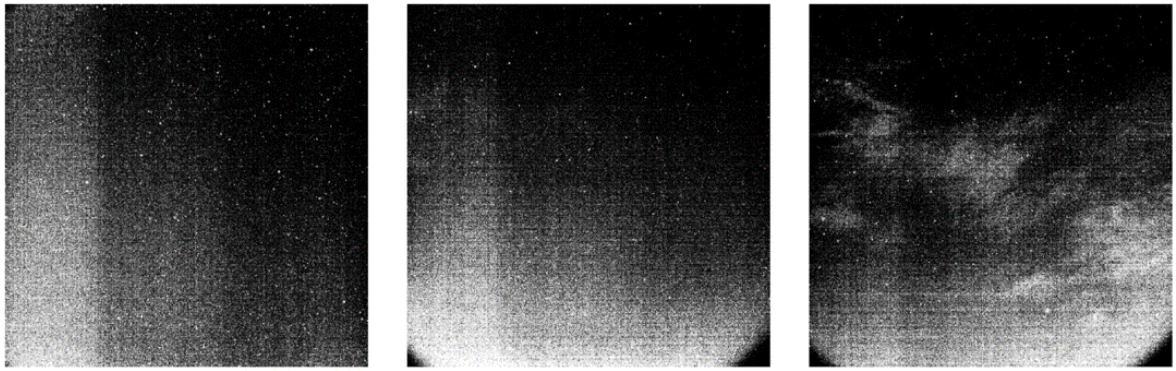
(a) (b) (c)

Figure 4. Centroiding results (a) star image with a big non-stellar bright object, (b) star position extraction by the first process, (c) star position extraction by the second process [26]

https://doi.org/10.31436/iiumej.v27i2.4246

Figure 4 provides a visual comparison of the results achieved by the proposed star centroiding algorithm, which includes steps to eliminate non-stellar, big-bright objects and extract star positions. The figures show the star number as the representation of the detected star blobs. Mahi et al. explicitly state that both proposed algorithms are limited in their ability to detect stars that are heavily affected by sunlight [26].

Zhao et al. (2024) proposed the Image Entropy-based Grayscale Iterative (IEGI) algorithm for extracting star spots from images affected by strong stray-light interference [9]. The method applies repeated grayscale iterations to suppress background noise based on local contrast information, followed by an object-to-pixel extraction stage that uses inner-outer templates and local image entropy to distinguish star regions from the background.



(a) (b) (c)

Figure 5. Real star image captured with the camera. (a) stray light condition 1; (b) stray light condition 2; (c) stray light condition 3 [9]

The study reported that the algorithm improved the star-detection rate by more than 45% compared to conventional max–mean filtering methods, while also reducing false detections by more than 50% relative to the Top-Hat transform. Experimental results with real star images showed that the method detected up to 93 star spots under severe interference conditions, where several conventional techniques failed. However, the iterative processing stages increase computational complexity, resulting in longer processing times of 0.69-1.04 seconds.

Table 2. Qualitative comparison of the related works

Method	Complexity	Computation Efficiency	Accuracy
SSA [38]	Moderate due to pattern-matching steps	Moderate but high in processing time	Good under moderate noise conditions
IEGI [9]	High	Low, it has a long running time due to the multiple iterations	High
1D CNN [8]	Low to Moderate	High with the use of sparse vectors allows skipping zero elements during convolution	High
Proposed	Moderate	Moderate to high	Improved robustness with lower complexity

Table 2 presents a qualitative comparison of several centroiding and star-detection methods in terms of computational complexity, efficiency, and detection accuracy. Based on the comparison table, more advanced methods, such as iterative weighting and entropy-based approaches, generally achieve higher accuracy and robustness, but they introduce greater computational complexity and longer processing times. The proposed method aims to provide

a balance between computational simplicity and robust centroiding performance under low SNR lunar imaging conditions.

3. METHODOLOGY

The proposed centroiding pipeline consists of three sequential stages: (1) image preprocessing for noise suppression, (2) binary segmentation of candidate star regions, and (3) intensity-weighted centroid extraction. The complete workflow is illustrated in Figure 6. The algorithm was implemented in MATLAB R2025b on a workstation with an Intel Core i9-13900K CPU (32 cores), 64 GB DDR4 RAM, and Windows 11 Pro. The implementation relies on functions from the Image Processing Toolbox (`medfilt2`, `imbinarize`, `regionprops`, `bwlabel`) and the Computer Vision Toolbox; no GPU acceleration is required, which is consistent with the target deployment on resource-constrained onboard processors. To facilitate replication, all source images, generated noisy datasets, and processing scripts were organized in a single MATLAB project, with each noise model encapsulated in a dedicated function (`addGaussianNoise`, `addPoissonNoise`, `addSaltPepper`, `addSpeckle`, `addSolarGlare`).

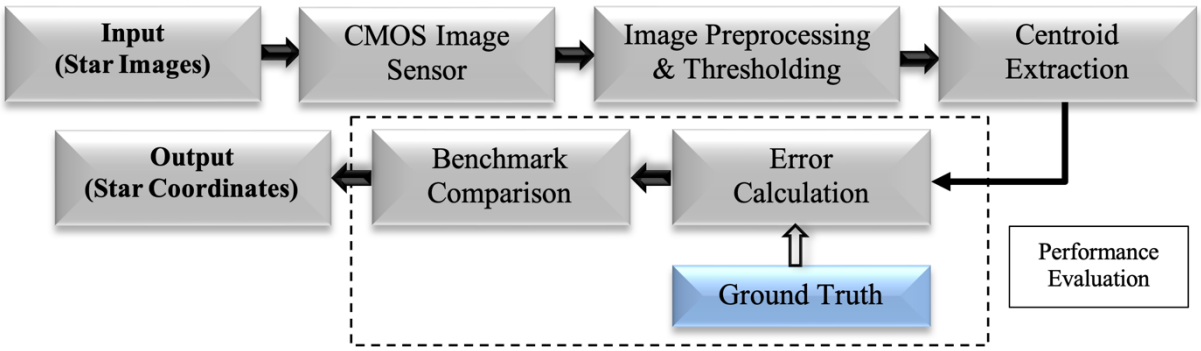


Figure 6. Block diagram of the proposed algorithm

A dataset of 30 star-field images was generated from the Stellarium Web online sky simulator [39]. Each image was exported as a PNG at 1920×1080 pixels (300 dpi) to emulate the field-of-view of a typical CMOS-based star camera. Camera calibration was excluded from the present study because consistent, clear-sky observations under local environmental conditions are difficult to obtain; this constraint is acknowledged in Section 5 as a direction for future work. Each clean image was then independently corrupted by five noise models (Gaussian, Poisson, salt-and-pepper, speckle, and a synthetic solar-glare gradient), producing a degraded dataset with measured SNRs ranging from −4.71 dB to 0.72 dB. The processing chain (median filter → threshold → centroid) was then applied to every image. Detected centroids were compared with the ground-truth coordinates extracted from the Stellarium star catalog, and three quantitative metrics were computed for each image: per-star Euclidean distance, root-mean-square error (RMSE), and false detection rate (FDR), as defined in Section 4.

3.1. Background Removal

A 3×3 median filter, implemented in MATLAB via the `medfilt2` function with symmetric boundary padding, is applied as the first preprocessing stage to suppress impulse-like noise without blurring star cores [40]. Within each 3×3 neighborhood, the central pixel is replaced by the median intensity, which preserves stellar peaks while removing isolated bright

https://doi.org/10.31436/iiumej.v27i2.4246

outliers caused by cosmic-ray hits, hot pixels, and salt-and-pepper sensor noise [41]. The 3×3 kernel size was selected after pilot tests with 3×3, 5×5, and 7×7 kernels: larger kernels were observed to over-smooth dim stellar peaks and shift their estimated locations, whereas the 3×3 kernel preserved sub-pixel peak structure while still effectively suppressing impulse noise. This sub-pixel preservation is critical because excessive smoothing biases the downstream centroid estimate.

As an example, consider a 3×3 pixel:

$$\begin{bmatrix} 34 & 36 & 40 \\ 38 & 255 & 41 \\ 35 & 39 & 37 \end{bmatrix}$$

The sorted values in the window are:

$$\{34, 35, 36, 37, \mathbf{38}, 39, 40, 41, 255\}$$

The median value is 38, which replaces the center pixel. The converted image becomes:

$$\begin{bmatrix} 34 & 36 & 40 \\ 38 & \mathbf{38} & 41 \\ 35 & 39 & 37 \end{bmatrix}$$

This method is then applied to the entire image to help produce cleaner star regions with preserved edges. This will allow subsequent detection and centroiding steps to operate more reliably under noisy imaging conditions.

3.2. Binary Segmentation

After median filtering, the denoised image is converted into a binary mask to isolate candidate stellar regions. A simple global thresholding scheme was selected over locally-adaptive methods (e.g., Otsu, Sauvola) because the lunar use case prioritizes deterministic, low-latency execution on embedded hardware [42], [43]. This segmentation is performed using a global threshold, T, which is defined as a fraction of the maximum intensity value in the image:

$$T = \alpha \cdot \max(I) \tag{1}$$

where I represents the pixel intensity, while α is a threshold factor ranging from 0 to 1 that controls the segmentation's sensitivity. The cutoff value T is a threshold factor in the range [0, 1] that controls segmentation sensitivity. To select an appropriate value, a sweep was performed over $\alpha \in \{0.2, 0.25, 0.3, 0.35, 0.4\}$ on the training subset. Values below 0.25 produced excessive false positives by retaining residual background noise, while values above 0.35 caused dim stars to be eliminated entirely. The setting $\alpha = 0.30$ minimized FDR without compromising recall and was therefore adopted for all subsequent experiments. Any pixel whose denoised intensity exceeds 30% of the per-image peak intensity is classified as a foreground (candidate-star) pixel; otherwise, it is classified as background. This binarization is expressed as:

$$B(x,y) = \begin{cases} 1, & if\ I(x,y) > T \\ 0, & otherwise \end{cases} \tag{2}$$

https://doi.org/10.31436/iiumej.v27i2.4246

In this formulation, $I(x, y)$ denotes the intensity of the pixel at position (x, y), and $B(x, y)$ is the resulting binary image. Foreground pixels $(B = 1)$ correspond to candidate star regions, whereas background pixels $(B = 0)$ represent noise and dark sky areas.

3.3. Centroid Extraction

Connected foreground regions in the binary mask are then labeled in MATLAB using the `bwlabel` function (8-connectivity) and filtered by area (minimum 3 pixels, maximum 200 pixels) to discard noise blobs and saturated artifacts. For each retained region, an intensity-weighted centroid is computed. In real star images, the captured intensity is non-uniform: brightness peaks near the stellar core and decays outward following an approximately Gaussian point-spread function. A simple binary centroid (geometric mean of mask pixels), therefore, introduces bias under noise, since it treats noisy boundary pixels and bright core pixels equally.

To address this, the COM formulation is applied, where pixel contributions are proportional to their intensities [44]. Concretely, the original (denoised) intensity image, the binary mask, and the labeled region map are passed to a custom MATLAB routine that, for each connected component, accumulates the intensity-weighted moments over the bounding box of the region; this gives a more stable, sub-pixel centroid estimate than equal weighting under low-SNR conditions.

The centroid coordinates are calculated as in Eq. (3):

$$x_c = \frac{\sum_{x,y} x \cdot I(x,y) \cdot w(x,y)}{\sum_{x,y} I(x,y) \cdot w(x,y)}, y_c = \frac{\sum_{x,y} y \cdot I(x,y) \cdot w(x,y)}{\sum_{x,y} I(x,y) \cdot w(x,y)} \quad (3)$$

where $I(x, y)$ is the pixel intensity at the coordinate (x, y), and the summation is taken over all pixels within the labeled star region.

4. Results and Discussion

This section reports the experimental evaluation of the proposed centroiding algorithm against three baselines (standard COM, Gaussian fitting, and SSA) on the 30-image dataset described in Section 3. Section 4.1 first characterizes the noise content and SNR distribution of the dataset; Section 4.2 then presents per-image and aggregate accuracy comparisons, supported by a critical discussion of failure modes and the practical implications of the observed performance gap.

4.1. Low SNR Star Images

Each clean image was independently corrupted by the five noise models defined in Table 1: Gaussian, Poisson, salt-and-pepper, speckle, and a synthetic solar-reflection (glare) gradient. SNR values were measured per image to characterize the severity of degradation, and the dataset was constructed to span mild-to-heavy noise conditions. Figure 7 shows the noise-free baseline used in all simulations.

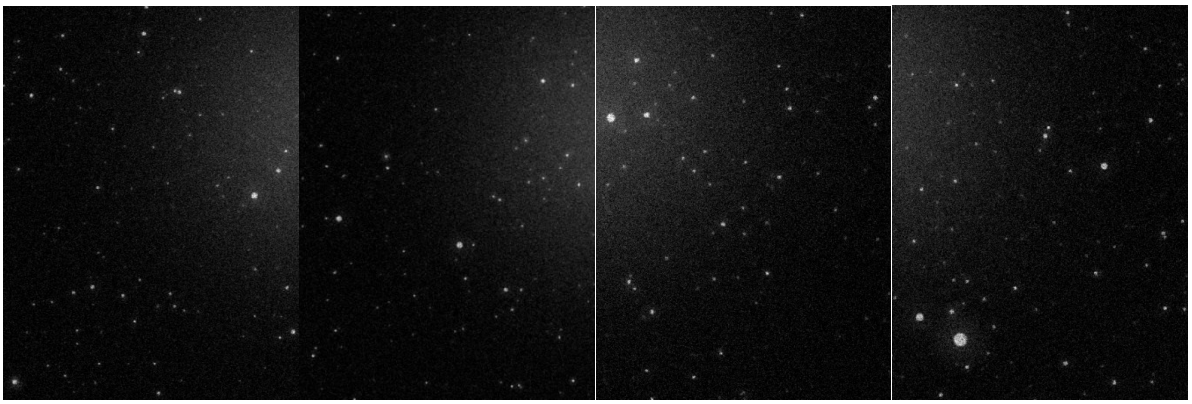
The image has a 16:9 landscape layout and is an HD image with a resolution of 1920×1080 pixels at 300 dpi. This consistent image specification is maintained across all evaluations to ensure fair comparison and to obtain clearer, more reliable results.

https://doi.org/10.31436/iiumej.v27i2.4246



Figure 7. Sample of clean and actual size of the star image

Figure 8(a)-(d) shows the selected data of 4 images, and Table 3 presents the SNR values for each image. For clarity, each figure has been cropped and zoomed from the original test images to allow detailed visual inspection of the star blobs and their distortions under the different noise conditions. The actual image sizes are shown in Figure 6.



(a) Image 1 (b) Image 2 (c) Image 3 (d) Image 4

Figure 8. Selected star images in the dataset

Table 3. SNR of selected star images from the dataset

Image	**1**	**2**	**3**	**4**
SNR (dB)	-4.71	-2.09	0.72	0.53

In such conditions, the reflected starlight captured by a star sensor is often weak due to the absence of an atmosphere, harsh illumination contrast, and scattered light from the lunar regolith [11], [45], [46]. Therefore, testing under these SNR levels provides a realistic performance evaluation of the centroiding algorithm.

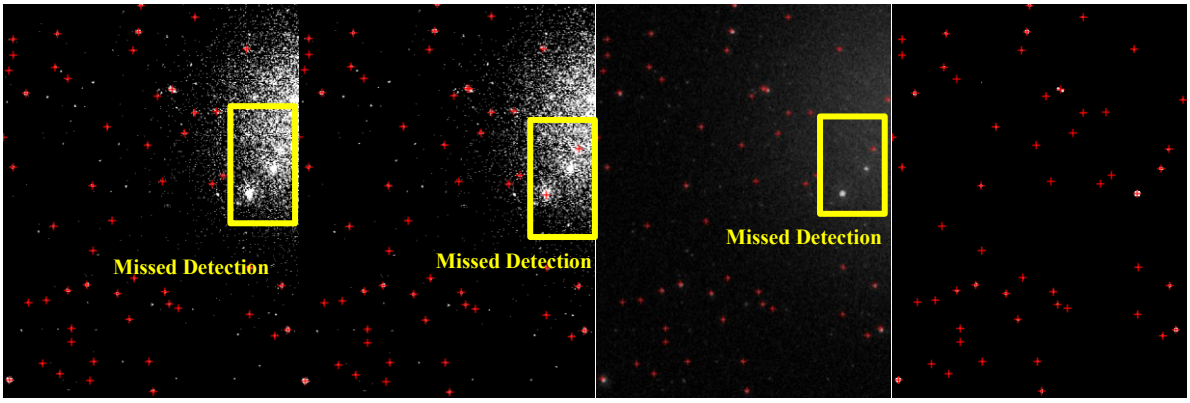
https://doi.org/10.31436/iiumej.v27i2.4246

4.2. Accuracy Analysis and Comparison

The proposed centroiding method is benchmarked against three established baselines, standard Center of Mass (COM), Gaussian fitting, and Sieve Search Algorithm (SSA), selected to span the design spectrum from purely intensity-weighted (COM) to model-fitting (Gaussian) to sub-region-search (SSA) approaches. Per-star localization accuracy is quantified using the Euclidean distance error (Eq. 4), which measures the straight-line distance between the detected centroid and its ground-truth position in pixel coordinates [47].

$$d = \sqrt{\left(x_i - x_i^{GT}\right)^2 + \left(y_i - y_i^{GT}\right)^2} \tag{4}$$

where GT is the coordinate of the ground truth. The comparison of position error or Euclidean distance is shown in graphs in Figures 9, 11, 13, and 15. The x-axis refers to the star count, where the count starts from the leftmost side of the star image until the rightmost side of the image. For reference, the actual star image size for the star count is shown in Figure 7.



(a) Image 1 (b) Image 2 (c) Image 3 (d) Image 4

Figure 9. Centroiding results for Image 1 under an SNR of -4.71 dB. (a) The COM method shows several missed detections in dense star regions due to noise interference within the highlighted area. (b) Gaussian Fitting exhibits slightly improved centroid placement; however, some weak stars are still undetected under low SNR conditions. (c) SSA demonstrates better noise suppression yet misses several faint stars. (d) The proposed method successfully detects and locates more star centroids within the same region.

Figure 10 quantifies the visual differences observed in Figure 9. SSA exhibits the largest Euclidean distance among all methods, with errors exceeding 4 pixels in several stars. This is attributable to its sub-region search mechanism, which becomes unstable when the local intensity profile is dominated by noise rather than a clear stellar peak. All four methods show elevated errors on the right side of the frame (approaching 5 pixels) along with several missed detections; this region coincides with the highest local noise concentration in Image 1. The proposed algorithm, by combining median pre-filtering with intensity-weighted COM, retains most centroid errors below 1 pixel, indicating that the early impulse-noise suppression step protects the downstream centroid estimator from being skewed by spurious bright pixels.

https://doi.org/10.31436/iiumej.v27i2.4246

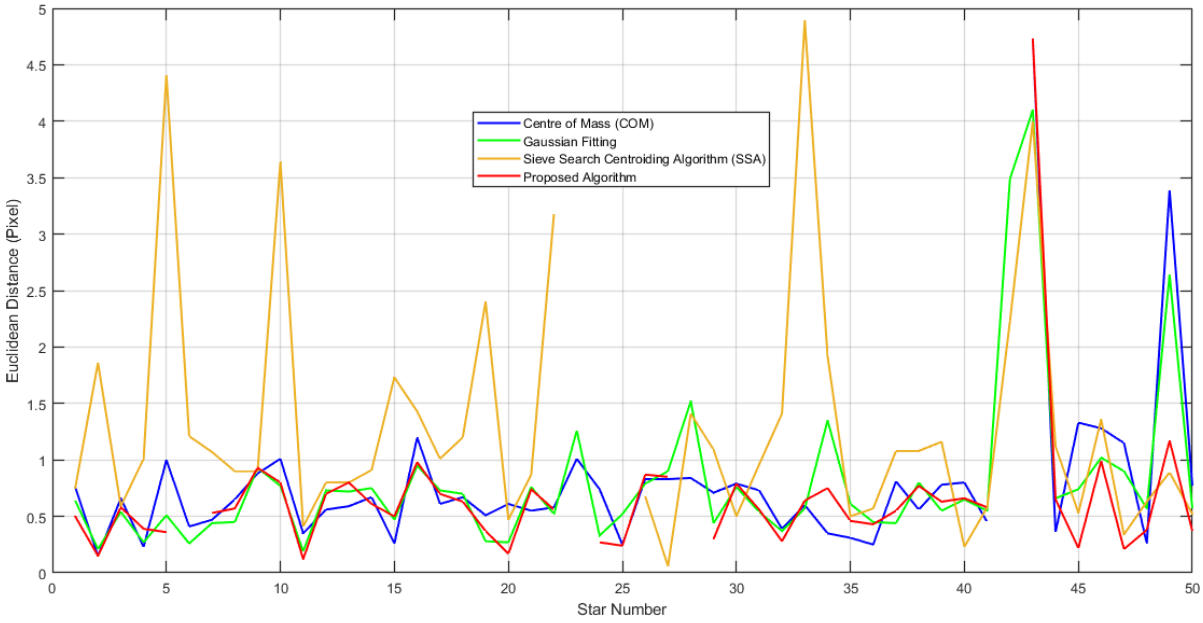
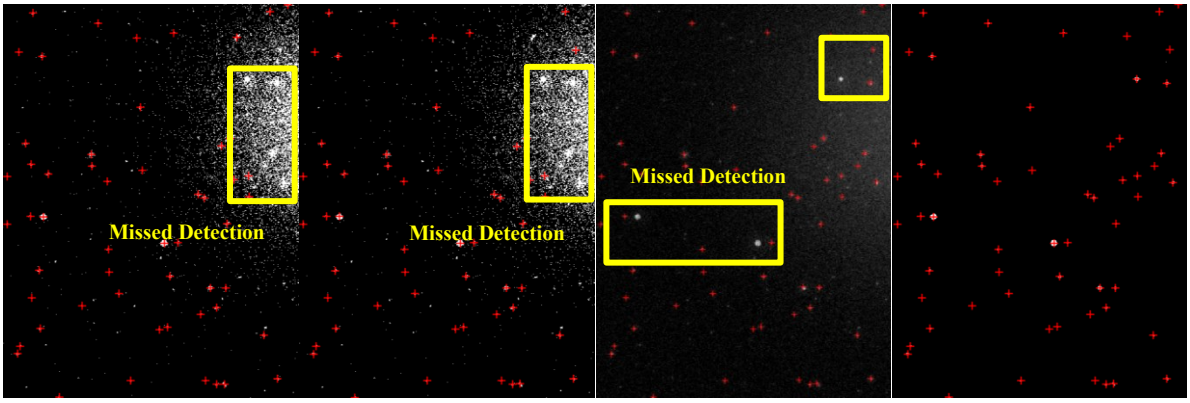


Figure 10. Comparison of Euclidean distance for each star in Image 1



(a) Image 1 (b) Image 2 (c) Image 3 (d) Image 4

Figure 11. Centroiding results for Image 2 under an SNR of -2.09 dB. (a) The COM method shows poor detection in high-noise regions, with several missed centroids in the highlighted upper area. (b) Gaussian Fitting provides slightly better accuracy but still fails to detect faint stars affected by noise clustering. (c) SSA improves star localization, but multiple faint stars in both the upper and middle regions remain undetected. (d) The proposed method achieves the most stable detection by maintaining accurate centroid detection even in low-intensity regions.

Figure 12 shows that, at the marginally higher SNR of Image 2 (−2.09 dB), all methods produce errors in the 1–2 pixel range, but with markedly different reliability. SSA again records the largest positional errors and the highest number of missed stars, consistent with its known sensitivity to parameter choices when the local SNR is low. The proposed method, COM, and Gaussian fitting are within 0.5 pixels of each other for detected stars, but the proposed method detects the most true stars, indicating that the error reduction is achieved without sacrificing recall. This is a non-trivial property: many noise-suppression schemes improve precision by raising the detection threshold, which trades recall for accuracy.

https://doi.org/10.31436/iiumej.v27i2.4246

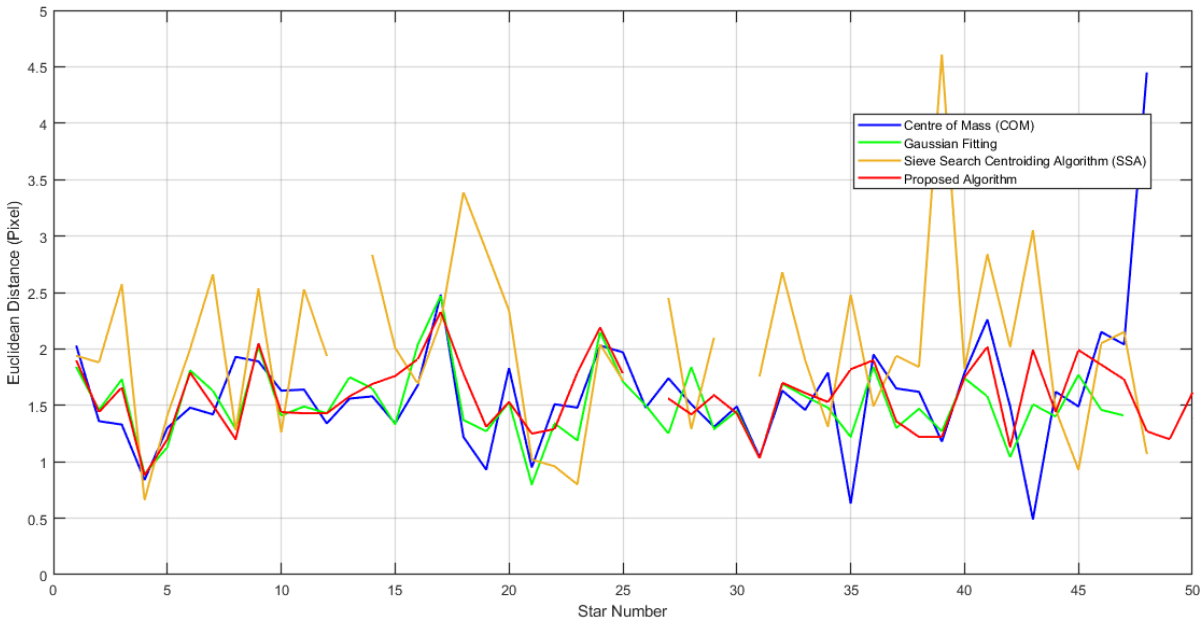
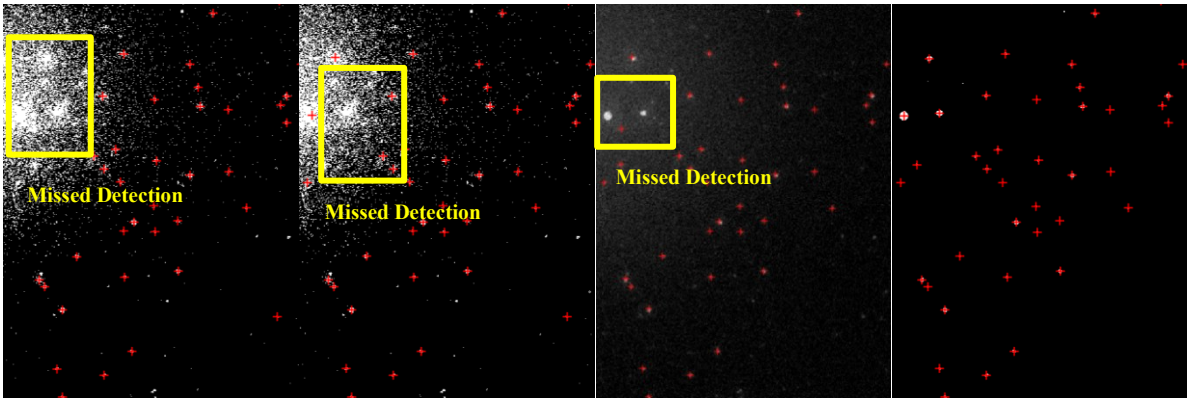


Figure 12. Comparison of Euclidean distance for each star in Image 2



(a) Image 1 (b) Image 2 (c) Image 3 (d) Image 4

Figure 13. Centroiding results for Image 3 under an SNR of 0.72 dB. (a) The COM method exhibits centroid deviation in the upper-left region due to noise, leading to multiple missed detections. (b) Gaussian Fitting provides slightly improved localization but still fails to resolve faint stars affected by glare from sunlight. (c) SSA enhances centroid clarity but misses several faint stars within the noisy region. (d) The proposed method successfully detects the majority of visible stars with precise centroid placement, outperforming the other methods in maintaining centroid accuracy under moderate noise levels.

Image 3 introduces a spatially heterogeneous noise distribution, with degradation concentrated on the left side of the frame. Figure 14 mirrors this asymmetry: errors and missed detections cluster on the left and decrease monotonically toward the right. This pattern is informative because it isolates the methods' behavior in localized high-noise regions independent of the global SNR. The proposed algorithm keeps most centroid errors below 2 pixels even within the heavily corrupted region, while the baselines show abrupt error spikes. This robustness to spatially non-stationary noise is particularly relevant for lunar imaging, where solar-glare gradients produce exactly this kind of one-sided degradation.

https://doi.org/10.31436/iiumej.v27i2.4246

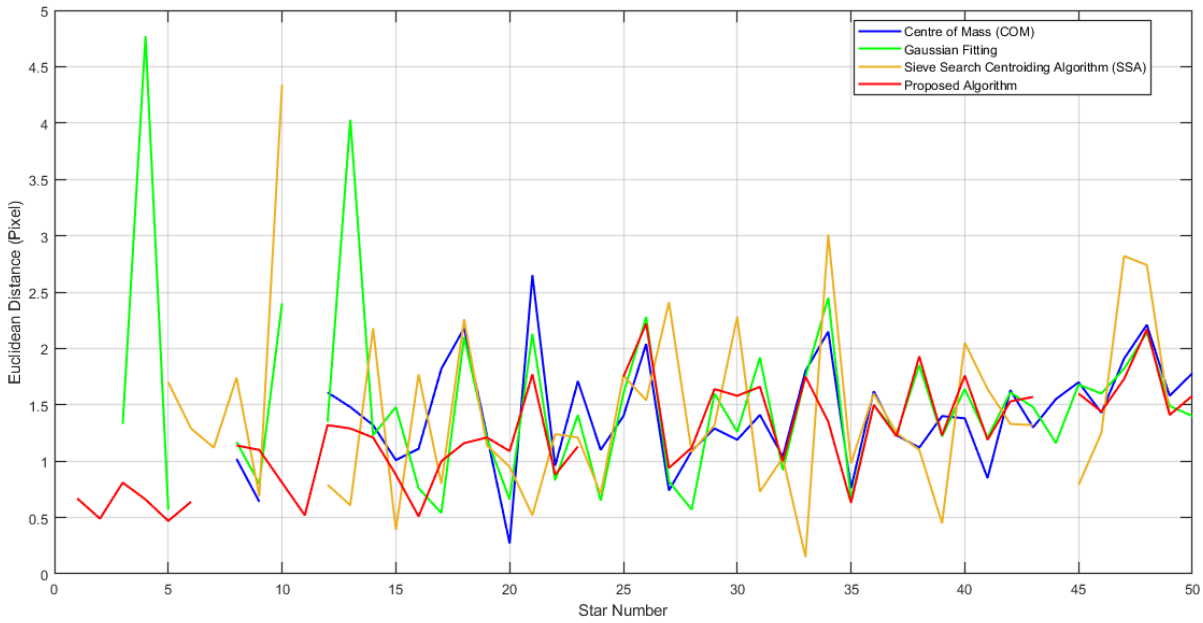


Figure 14. Comparison of Euclidean distance for each star in Image 3

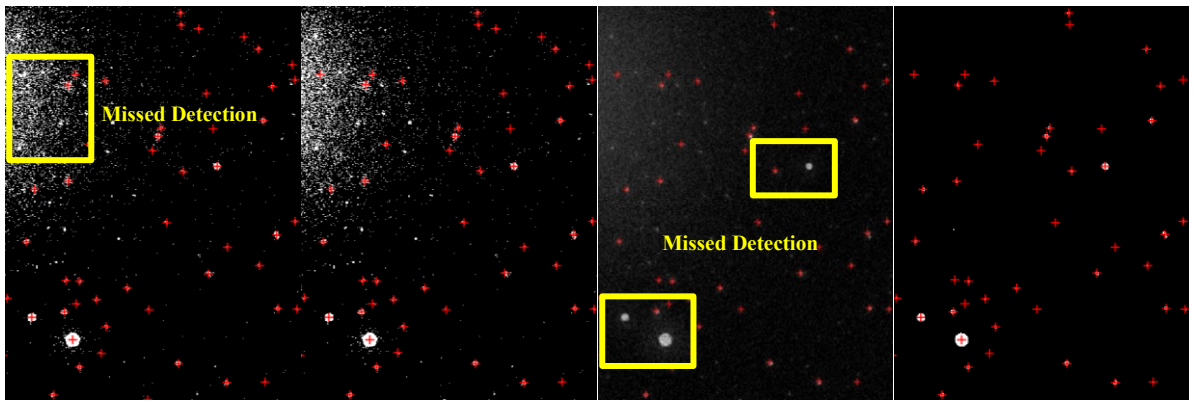


Figure 15. Centroiding results for Image 4 under an SNR of 0.53 dB. (a) The COM method provides good centroid placement with fewer missed detections; however, a few faint stars remain undetected near the noisy region. (b) The Gaussian Fitting method yields more stable centroiding than COM. (c) SSA exhibits multiple missed detections in a few regions, especially for faint stars. (d) The proposed algorithm effectively identifies most stars across the frame with minimal missed detections, demonstrating high robustness and accuracy in low-SNR imaging environments.

Figure 16 corresponds to Image 4, the most challenging case in the qualitative subset. All baselines exhibit position errors of approximately 4 pixels, which is an order of magnitude beyond the sub-pixel accuracy required for arc-second attitude determination. SSA exhibits a missed-detection pattern similar to Figure 13(c), suggesting that its sub-region search degrades systematically rather than gracefully under speckle and glare. Both Gaussian fitting and COM produce centroid errors that exceed those of the proposed method by 30–50%. Notably, the proposed algorithm's error distribution remains tight in this regime, indicating that the median-filter + thresholding combination effectively decouples the centroid estimator from the worst-case noise tail, rather than merely shifting the error distribution.

https://doi.org/10.31436/iiumej.v27i2.4246

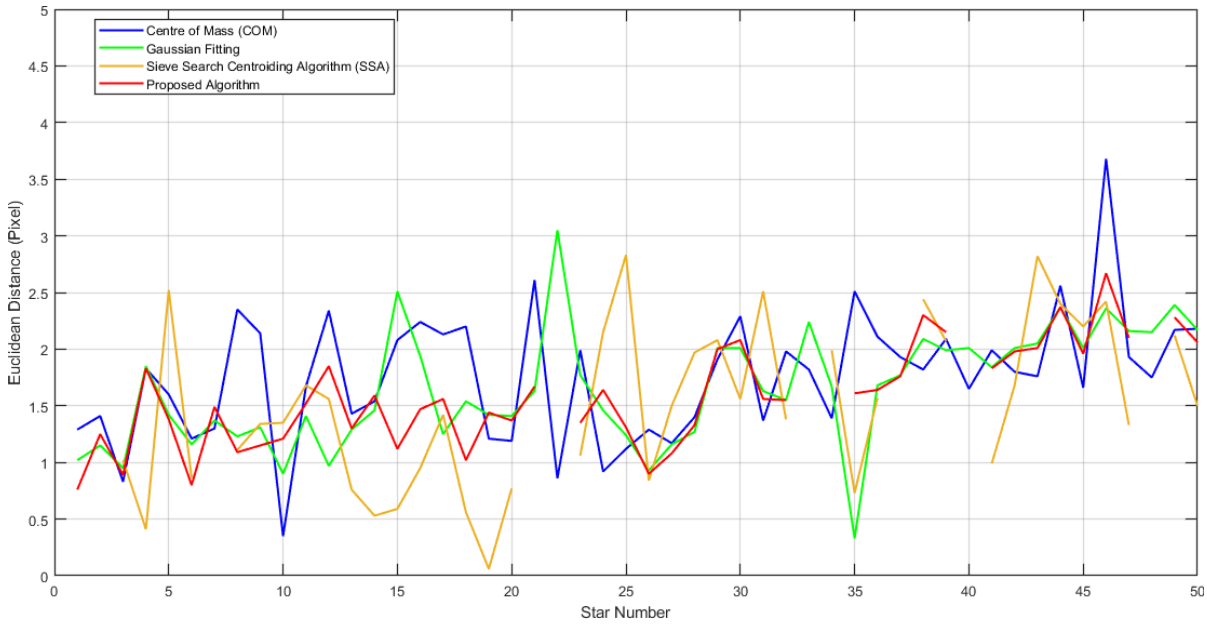
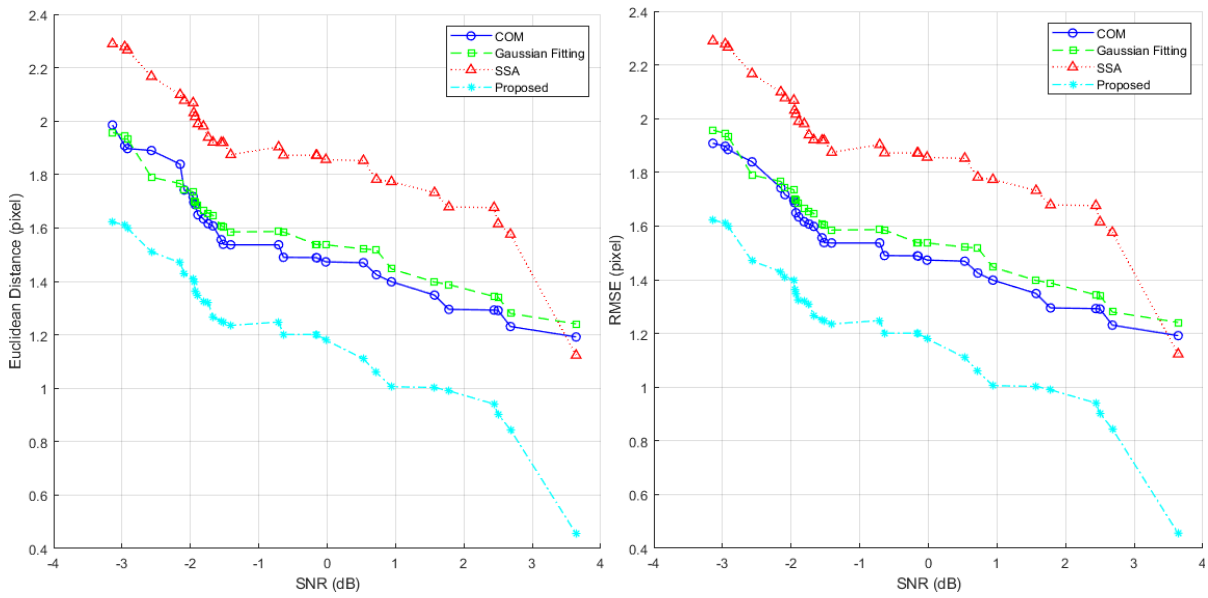


Figure 16. Comparison of Euclidean distance for each star in Image 4

To summarize the overall error performance for the star images, $RMSE$ is used to evaluate a statistical overview of error magnitude across all detected stars, as in Eq. (5) [48]:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{5}$$

where $e_i = y_i - \hat{y}_i$ is the Euclidean error that has been calculated, and N is the number of stars. This experiment is also extended to the remaining 26 star field images to obtain a more comprehensive set of results for evaluation. The performance results for all methods on every 30 images are shown in Figure 17.



(a) Euclidean Distance (b) RMSE

Figure 17. Performance comparison between existing methods and the proposed method

Aggregating across all 30 images (Figure 17) reveals a consistent ranking: SSA records the highest position error, followed by Gaussian fitting, then standard COM, with the proposed method producing the lowest values on both Euclidean distance and RMSE axes. Importantly, the gap between the proposed method and the next-best baseline widens as the average noise level increases, indicating that the advantage is most pronounced precisely in the operating regime that lunar missions encounter. To complement positional accuracy, detection reliability is quantified using the false detection rate (FDR), which is critical because false centroids contaminate the subsequent star-identification stage of a star tracker. A higher FDR indicates over-sensitivity to noise, while a lower FDR reflects more reliable detection [27]. The FDR can be expressed as in Eq. (6):

$$FDR = \frac{N_{false}}{N_{detected}} \times 100\% \quad (6)$$

Table 4 presents the quantitative comparison of the average centroiding accuracy and reliability among the four tested methods.

Table 4. Summary of the accuracy analysis of the average of the proposed method with other methods

Methods	RMSE	Euclidean Distance	False Detection Rate (%)
Center of Mass (COM)	1.551	1.373	13.949
Gaussian Fitting	1.599	1.368	9.983
Sieve Search Algorithm (SSA)	1.904	1.679	12.637
Proposed Algorithm	1.218	1.143	6.716

Three observations emerge from Table 4. First, the proposed method achieves the lowest values on every metric simultaneously: RMSE = 1.218 pixels (a 21.5% reduction over the best baseline, COM at 1.551), Euclidean distance error = 1.143 pixels (a 16.4% reduction over Gaussian fitting at 1.368), and FDR = 6.716% (a 32.7% reduction over Gaussian fitting at 9.983%). Joint dominance in both accuracy and reliability is uncommon in the literature, where most low-complexity methods improve one metric at the expense of the other. Second, SSA underperforms standard COM on every metric in this evaluation, despite its design objective of improved noise robustness. This is consistent with prior reports that SSA's sub-region partitioning is sensitive to parameter selection and degrades on noise-dominated, sparse star fields, which differ from the dense fields for which it was originally tuned. Third, Gaussian fitting yields a competitive Euclidean distance (1.368 pixels, second-best) but is penalized by a higher RMSE (1.599 pixels), indicating that its error distribution has a heavier tail; that is, when it fails, it fails badly. The proposed method's tighter RMSE–Euclidean gap suggests a more uniform error distribution and therefore more predictable behavior, a desirable property for safety-critical onboard navigation. Practically, the observed FDR of 6.716% means that fewer than 1 in 14 detected centroids is spurious, which is within the tolerance of standard star-identification algorithms (typically robust to FDR below ~10%).

Taken together, the per-image and aggregate results consistently show that the proposed pipeline delivers a favorable accuracy–reliability–complexity trade-off. The improvement is not driven by a single metric or a single noise model, but is observed across all five degradation types and the full SNR range tested. Two limitations of this study should, however, be acknowledged. First, the noise models are statistical surrogates for lunar imaging conditions; physics-based models that incorporate the camera point-spread function (PSF), pixel response

non-uniformity (PRNU), and dark-current shot noise would yield more realistic error estimates. Second, runtime profiling of representative spaceflight processors (e.g., FPGAs, DSPs) was outside the scope of this paper. Both limitations directly motivate the future work outlined in Section 5.

5. CONCLUSION

This paper presented a low-complexity yet robust star centroiding algorithm tailored to the demanding noise environment of lunar surface navigation. The pipeline integrates three deliberately simple components — 3×3 median filtering for impulse-noise suppression, an adaptive global threshold fixed at 30% of peak intensity for star segmentation, and an intensity-weighted center-of-mass estimator for sub-pixel localization — into a fully MATLAB-replicable workflow. Validated on 30 Stellarium-derived star fields corrupted by Gaussian, Poisson, salt-and-pepper, speckle, and synthetic solar-glare noise, the proposed method outperformed three established baselines (standard COM, Gaussian fitting, and SSA) on every aggregate metric: it achieved an average RMSE of 1.218 pixels (a 21.5% reduction over the best baseline), an average Euclidean distance error of 1.143 pixels (a 16.4% reduction), and a false detection rate of 6.716% (a 32.7% reduction). Crucially, this dominance was consistent across all noise models and all SNR levels tested, indicating that the gains arise from the pipeline's architecture rather than from tuning to a specific scenario. The algorithm therefore advances the state of the art on the accuracy–reliability–complexity Pareto frontier and is well-suited to the resource-constrained processors typical of lunar landers, rovers, and CubeSat-class spacecraft.

Four directions for future work are identified. First, hardware deployment and runtime characterization on Field-Programmable Gate Arrays (FPGAs), Digital Signal Processors (DSPs), and CubeSat-class edge platforms will quantify the algorithm's achievable frame rate, power envelope, and timing determinism, complementing the present accuracy-only evaluation. Second, the statistical noise models used here will be augmented with physics-based degradation models that incorporate the optical point spread function (PSF), CMOS pixel response non-uniformity (PRNU), dark-current shot noise, and rad-hardened sensor effects to better predict in-flight performance. Third, an end-to-end star-camera calibration pipeline will be developed to enable real-time star acquisition under real night-sky conditions and to support a Lost-in-Space (LIS) operating mode in which the system autonomously identifies stars without prior attitude information. Fourth, the pipeline will be evaluated on representative lunar mission imagery (e.g., Chang'e and Yutu-2 datasets) to assess generalization beyond the simulated benchmarks reported here. Together, these steps will progress the proposed algorithm from a validated software prototype toward a deployable component of an autonomous lunar navigation system.

ACKNOWLEDGMENT

This research was supported by the Ministry of Higher Education of Malaysia (MOHE) Fundamental Research Grant Scheme (FRGS) code number FRGS/1/2024/TK07/UIAM/02/1 and supported by KHAIR Award Fund from the Center for Postgraduate Studies (CPS) IIUM.

REFERENCES

[1] F. Wu *et al.*, "Anti-noise Star Image Extraction Algorithm for Star Trackers Based on YOLOv5," in *2023 Intelligent Methods, Systems, and Applications (IMSA)*, 2023, pp. 149–154. doi: 10.1109/IMSA58542.2023.10217394.

[2] Anqi Fan, "China-russia lunar base collaboration 'A perfect match.' ," Global Times.

[3] B. Yağlioğlu, B. B. İnaltekin, and M. Gökten, "The First Turkish Lunar Mission Part 1: Programmatic, Mission and System Aspects," in *2023 10th International Conference on Recent Advances in Air and Space Technologies (RAST)*, IEEE, Jun. 2023, pp. 1–6. doi: 10.1109/RAST57548.2023.10197907.

[4] B. B. İnaltekin, B. Yağlioğlu, and M. Ceylan, "The First Turkish Lunar Mission Part 2: Overview of Space Segment," in *2023 10th International Conference on Recent Advances in Air and Space Technologies (RAST)*, IEEE, Jun. 2023, pp. 1–5. doi: 10.1109/RAST57548.2023.10197853.

[5] M. B. Trageser and D. G. Hoag, "Apollo Spacecraft Guidance System," *IFAC Proceedings Volumes*, vol. 2, no. 1, pp. 435–463, Jun. 1965, doi: 10.1016/S1474-6670(17)69119-5.

[6] C. Li *et al.*, "Overview of the Chang'e-4 Mission: Opening the Frontier of Scientific Exploration of the Lunar Far Side," *Space Sci. Rev.*, vol. 217, no. 2, p. 35, Mar. 2021, doi: 10.1007/s11214-021-00793-z.

[7] Q. Wang and J. Liu, "A Chang'e-4 mission concept and vision of future Chinese lunar exploration activities," *Acta Astronaut.*, vol. 127, pp. 678–683, Oct. 2016, doi: 10.1016/j.actaastro.2016.06.024.

[8] B. Wang, H. Wang, and Z. Jin, "An Efficient and Robust Star Identification Algorithm Based on Neural Networks," *Sensors*, vol. 21, no. 22, p. 7686, Nov. 2021, doi: 10.3390/s21227686.

[9] Q. Zhao, J. Liao, D. Zhang, and J. Feng, "Grayscale Iterative Star Spot Extraction Algorithm Based on Image Entropy," *Applied Sciences*, vol. 14, no. 20, p. 9207, Oct. 2024, doi: 10.3390/app14209207.

[10] D. E. Wilhelms, J. F. McCauley, and N. J. Trask, "The geologic history of the Moon," 1987. doi: 10.3133/pp1348.

[11] H. Lin *et al.*, "Photometric properties of lunar regolith revealed by the Yutu-2 rover," *Astron. Astrophys.*, vol. 638, p. A35, Jun. 2020, doi: 10.1051/0004-6361/202037859.

[12] R. Li *et al.*, "Cone penetration resistance of CUMT-1 lunar regolith simulant under magnetic-similitude lunar gravity condition," *Acta Geotech.*, vol. 18, no. 12, pp. 6725–6744, Dec. 2023, doi: 10.1007/s11440-023-01971-9.

[13] NASA, "Moon Composition & Structure." Accessed: Oct. 03, 2025. [Online]. Available: https://science.nasa.gov/moon/composition/

[14] Y.-L. Zou, J.-Z. Liu, J.-J. Liu, and T. Xu, "Reflectance Spectral Characteristics of Lunar Surface Materials," *Chinese Journal of Astronomy and Astrophysics*, vol. 4, no. 1, pp. 97–104, Feb. 2004, doi: 10.1088/1009-9271/4/1/97.

[15] M. Horányi, J. R. Szalay, and X. Wang, "The lunar dust environment: concerns for Moon-based astronomy," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 382, no. 2271, May 2024, doi: 10.1098/rsta.2023.0075.

[16] J. J. Rennilson and D. R. Criswell, "Surveyor observations of lunar horizon-glow," *The Moon*, vol. 10, no. 2, pp. 121–142, Jun. 1974, doi: 10.1007/BF00655715.

[17] NASA, "The Apollo Lunar Surface Journal and Apollo Flight Journal." Accessed: Nov. 23, 2025. [Online]. Available: https://www.nasa.gov/history/alsj-and-afj/

[18] Rafael Alanis, "Diviner Global Composition," NASA Science. Accessed: Dec. 04, 2025. [Online]. Available: https://science.nasa.gov/photojournal/diviner-global-composition/

[19] C. C. Liebe, "Accuracy performance of star trackers - a tutorial," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 2, pp. 587–599, Apr. 2002, doi: 10.1109/TAES.2002.1008988.

[20] F. N. Chukhovskii, P. V. Konarev, and V. V. Volkov, "Denoising of the Poisson-Noise Statistics 2D Image Patterns in the Computer X-ray Diffraction Tomography," *Crystals (Basel).*, vol. 13, no. 4, p. 561, Mar. 2023, doi: 10.3390/cryst13040561.

[21] R. K. Painam and M. Suchetha, "Comparative Performance Analysis of Spatial Domain Filters for Removing Speckle Noise in SAR images," in *2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*, IEEE, Feb. 2022, pp. 01–05. doi: 10.1109/ICEEICT53079.2022.9768585.

[22] S. V. Parhad, K. K. Warhade, and S. S. Shitole, "Speckle noise reduction in sar images using improved filtering and supervised classification," *Multimed. Tools Appl.*, vol. 83, no. 18, pp. 54615–54636, Dec. 2023, doi: 10.1007/s11042-023-17648-0.

[23] W. Xu, Y. Liang, W. Chen, and F. Wang, "Recent advances of stretched Gaussian distribution underlying Hausdorff fractal distance and its applications in fitting stretched Gaussian noise," *Physica A: Statistical Mechanics and its Applications*, vol. 539, Feb. 2020, doi: 10.1016/j.physa.2019.122996.

[24] O. O. Khalifa, *Digital Image Processing and Computer Vision : Fundamentals and Applications*. IIUM Press, 2021.

[25] Y. Fan, H. Xiao, W. Cao, L. Zuo, and S. Chen, "FPGA Implementation of Real-time Star Centroid Extraction Algorithm," in *2019 IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP)*, 2019, pp. 395–399. doi: 10.1109/ICICSP48821.2019.8958483.

[26] Z. Mahi, M. S. Karoui, and M. Keche, "A Non-Stellar Big Bright Object Elimination and Star Centroid Extraction Algorithm for Star Tracker," in *2024 IEEE Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS)*, 2024, pp. 1–5. doi: 10.1109/M2GARSS57310.2024.10537475.

[27] Z. Mahi and M. S. Karoui, "A Comparative Study of Star Detection Methods for a Satellite-Onboard Star Tracker," in *2019 International Conference on Advanced Electrical Engineering (ICAEE)*, 2019, pp. 1–6. doi: 10.1109/ICAEE47123.2019.9015160.

[28] M. Ceresoli, F. Maccari, E. Moresi, and M. Lavagna, "Design and station-keeping strategies for robust lunar navigation constellations," *Acta Astronaut.*, vol. 236, pp. 20–31, Nov. 2025, doi: 10.1016/j.actaastro.2025.06.025.

[29] F. Xing, Y. Dong, Z. You, and Q. Zhou, "APS star tracker and attitude estimation," in *2006 1st International Symposium on Systems and Control in Aerospace and Astronautics*, 2006, pp. 5 pp. – 38. doi: 10.1109/ISSCAA.2006.1627697.

[30] H. Cai, Z. Yang, X. Cao, W. Xia, and X. Xu, "A new iterative triclass thresholding technique in image segmentation," *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1038–1046, Mar. 2014, doi: 10.1109/TIP.2014.2298981.

[31] H. Liang, H. Jia, Z. Xing, J. Ma, and X. Peng, "Modified Grasshopper Algorithm-Based Multilevel Thresholding for Color Image Segmentation," *IEEE Access*, vol. 7, pp. 11258–11295, 2019, doi: 10.1109/ACCESS.2019.2891673.

[32] R. C. Gonzalez and R. E. Woods, *Digital Image Processing, 4th ed.*, 4th ed. Pearson, 2018.

[33] R. Raman, V. Kumar, B. G. Pillai, D. Rabadiya, S. Patre, and R. Meenakshi, "Brain Image Denoising and Segmentation Using Advanced Mixed Thresholding and Edge Preservation Analysis Technique," in *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)*, IEEE, Apr. 2024, pp. 1–5. doi: 10.1109/ICKECS61492.2024.10617270.

[34] A. H. Razaman, Y. A. Ahmad, T. S. Gunawan, and O. O. Khalifa, "The Noise Reduction Algorithm for Star Detection," *PERINTIS EJournal*, vol. 15, no. 2, pp. 37–46, 2025, doi: 10.5281/zenodo.1808125.

[35] C. Chen, Q. Chen, C. Gao, N. Zhang, X. Wang, and Y. Zhai, "Method of Blob detection based on radon transform," in *2018 Chinese Control And Decision Conference (CCDC)*, 2018, pp. 5762–5767. doi: 10.1109/CCDC.2018.8408138.

[36] J. R. Janesick, *Scientific Charge-Coupled Devices*. SPIE, 2001. doi: 10.1117/3.374903.

[37] A. H. Razaman, Y. A. Ahmad, O. O. Khalifa, N. Saidin, and A. W. Azman, "Review Paper on Centroiding Algorithm for Lunar Navigation," in *2025 21st IEEE International Colloquium on Signal Processing & Its Applications (CSPA)*, IEEE, Feb. 2025, pp. 89–93. doi: 10.1109/CSPA64953.2025.10933304.

[38] V. C. Karaparambil, N. S. Manjarekar, and P. M. Singru, "Sieve Search Centroiding Algorithm for Star Sensors," *Sensors*, vol. 23, no. 6, 2023, doi: 10.3390/s23063222.

[39] "Stellarium Web Online Star Map," 2025. [Online]. Available: https://stellarium-web.org/

[40] M.-H. Hung, J.-S. Pan, and C.-H. Hsieh, "A fast algorithm of temporal median filter for background subtraction," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 5, pp. 33–40, Jun. 2014.

[41] Patrice, "Patrice's Lectures: COMPSCI 373 Image Filtering [Course Presentation]," in *The University of Auckland*, 2016. Accessed: Apr. 26, 2025. [Online]. Available: https://www.cs.auckland.ac.nz/courss/compsci373s1c/PatricesLectures/

[42] Z. Mahi, M. S. Karoui, and M. Keche, "A new star detection algorithm based on the combination of thresholding and filtering technique," in *2022 7th International Conference on Image and Signal Processing and their Applications (ISPA)*, IEEE, May 2022, pp. 1–6. doi: 10.1109/ISPA54004.2022.9786030.

[43] R. Singh, C. Prabha, and S. Abdulla, "Optimized CNN framework for malaria detection using Otsu thresholding−based image segmentation," *Sci. Rep.*, vol. 15, no. 1, p. 40117, Nov. 2025, doi: 10.1038/s41598-025-23961-5.

[44] X. Qian, H. Yu, and S. Chen, "A Global-Shutter Centroiding Measurement CMOS Image Sensor With Star Region SNR Improvement for Star Trackers," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 8, pp. 1555–1562, 2016, doi: 10.1109/TCSVT.2015.2469091.

[45] R. M. Marshal, O. Rüsch, C. Wöhler, K. Wohlfarth, and S. Velichko, "Photometry of LROC NAC resolved rock-rich regions on the Moon," *Icarus*, vol. 394, p. 115419, Apr. 2023, doi: 10.1016/j.icarus.2022.115419.

[46] S. Velichko, V. Korokhin, Y. Shkuratov, V. Kaydash, Y. Surkov, and G. Videen, "Photometric analysis of the Luna spacecraft landing sites," *Planet. Space Sci.*, vol. 216, p. 105475, Jul. 2022, doi: 10.1016/j.pss.2022.105475.

[47] P.-E. Danielsson, "Euclidean distance mapping," *Computer Graphics and Image Processing*, vol. 14, no. 3, pp. 227–248, Nov. 1980, doi: 10.1016/0146-664X(80)90054-4.

[48] Akshita Chugh, "MAE, MSE, RMSE, Coefficient of Determination, Adjusted R Squared — Which Metric is Better?," Medium. Accessed: Sep. 19, 2025. [Online]. Available: https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e

https://doi.org/10.31436/iiumej.v27i2.4341

Efficiency-Aware Multi-Class Spinal Disorder Classification Using CBAM-Enhanced Lightweight CNNs with Dual-Branch Fusion

TEDDY SURYA GUNAWAN[1*], NURUL JANNAH[2], MIRA KARTIWI[3], NOREHA ABDUL MALIK[1]

[1]*Electrical and Computer Engineering Dept., International Islamic University Malaysia, Malaysia*
[2]*Mechatronics Engineering Dept., International Islamic University Malaysia, Malaysia*
[3]*Information Systems Dept., International Islamic University Malaysia, Malaysia*

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

(Received: 6 March 2026; Accepted: 25 April 2026; Published online: 10 May 2026)

ABSTRACT: Spinal X-rays are still often read through manual measurements, yet the patients who most need timely assessment cannot afford delay, inconsistency, or heavy computational pipelines. Motivated by this clinical tension, this study proposes an efficiency-aware deep learning framework for three-class spinal disorder classification that asks a practical question rarely centered in prior work: not only which model is most accurate, but which model is accurate enough, light enough, and fast enough to matter in real screening settings. Using a public dataset of 338 subjects, five lightweight backbones, CBAM-enhanced variants, and a dual-branch fusion model were evaluated through stratified 5-fold cross-validation under multiple balancing strategies, with performance measured by accuracy, precision, recall, F1-score, parameter count, FLOPs, model size, latency, and throughput. The results reveal an unexpected pattern: bigger models do not win. MobileNetV3Small delivers the strongest efficiency-performance balance, reaching an F1-score of 0.962 with only 1.0 million parameters, while the best overall result is achieved by the Fusion_MNv3_MNAS model under augmentation-only training, with an F1-score of 0.976. Ablation findings further show that attention and fusion are not universally beneficial, but become most effective when paired with sufficient data-driven regularization, and that fine-tuning about 30% of backbone parameters yields the most favorable adaptation. Taken together, these findings show that performance in spinal X-ray classification depends less on model size alone than on the fit between architecture and training strategy. The study therefore offers a concrete and clinically relevant message: lightweight, well-regularized models can match or surpass heavier alternatives while remaining more practical for scalable deployment.

ABSTRAK: Radiograf tulang belakang masih kerap dinilai melalui pengukuran manual, sedangkan pesakit yang memerlukan rawatan awal tidak dapat menanggung kelewatan, ketidakselarasan, atau kebergantungan pada sistem pengiraan yang berat. Berpunca pada masalah klinikal ini, kajian ini mengemukakan satu rangka kerja pembelajaran mendalam berpaksikan kecekapan bagi pengelasan tiga kelas gangguan tulang belakang, dengan menumpukan persoalan praktikal yang jarang diberi perhatian dalam kajian terdahulu, iaitu bukan pada model mana yang paling tepat, tetapi model mana yang cukup tepat, cukup ringan, dan cukup pantas bagi persekitaran saringan klinikal. Dengan menggunakan satu set data awam yang melibatkan 338 subjek, lima model asas ringan, varian yang dipertingkatkan dengan CBAM, serta model gabungan dwi-cabang telah dinilai melalui pengesahan silang berstrata lima lipatan di bawah beberapa strategi pengimbangan kelas, dengan prestasi diukur menggunakan ketepatan, kejituan, keboleh ingatan, skor F1, bilangan parameter, FLOPs, saiz model, kependaman, dan kadar pemprosesan. Dapatan kajian menunjukkan satu corak yang tidak dijangka, iaitu model yang lebih besar tidak semestinya memberi prestasi terbaik.

MobileNetV3Small memperlihatkan keseimbangan paling kukuh antara kecekapan dan prestasi dengan mencapai skor F1 sebanyak 0.962 hanya menggunakan 1.0 juta parameter, manakala prestasi keseluruhan terbaik dicapai oleh model Fusion_MNv3_MNAS di bawah latihan berasaskan augmentasi sahaja dengan skor F1 sebanyak 0.976. Analisis ablasi seterusnya menunjukkan bahawa mekanisme perhatian dan gabungan tidak sentiasa memberikan manfaat secara menyeluruh, sebaliknya menjadi paling berkesan apabila dipadankan dengan regularisasi berasaskan data yang mencukupi, dan penalaan halus sekitar 30% parameter rangka asas menghasilkan penyesuaian yang terbaik. Secara keseluruhan, dapatan ini menunjukkan bahawa prestasi dalam pengelasan sinar-X tulang belakang kurang bergantung pada saiz model semata-mata, sebaliknya lebih dipengaruhi oleh kesesuaian antara seni bina model dan strategi latihan. Oleh itu, kajian ini membawa mesej yang jelas dan signifikan dari sudut klinikal, iaitu model ringan yang diregularisasikan dengan baik mampu menandingi malah mengatasi model yang lebih berat, iaitu kekal lebih praktikal bagi pelaksanaan berskala dalam persekitaran saringan klinikal.

KEYWORDS: *Spinal Disorder Classification, Deep Learning, Lightweight CNN, Attention Mechanism (CBAM), Efficiency-Aware Evaluation*

1. INTRODUCTION

A spinal X-ray may look routine, but the decision drawn from it can shape the rest of a patient's life. Conditions such as scoliosis and spondylolisthesis often begin as subtle structural abnormalities on radiographs, yet have significant outcomes. When diagnosis is delayed, the consequences can extend far beyond posture or back pain, progressing into deformity, functional limitation, and more invasive treatment. Scoliosis is commonly identified through abnormal lateral curvature of the spine, whereas spondylolisthesis is marked by the forward displacement of one vertebra over another. In daily clinical practice, both conditions are still largely assessed through manual radiographic measurements, including Cobb angle estimation and vertebral slippage grading. These methods remain clinically valuable, but they are time-consuming, operator-dependent, and vulnerable to inter- and intra-observer variation, which may affect diagnostic consistency and treatment planning [1, 2]. The problem, therefore, is not merely whether spinal disorders can be seen on X-rays, but whether they can be assessed quickly, consistently, and at scale.

Recent progress in deep learning has made automated spinal image analysis increasingly feasible. Convolutional neural networks (CNNs), particularly when used through transfer learning, have shown a strong ability to extract discriminative patterns directly from radiographs and to support accurate classification and detection of spinal abnormalities [3, 4]. Prior studies have reported high performance on three-class spinal classification tasks [1], while others have shown that model selection and optimization strategy can substantially affect the final outcome [5]. At the same time, related work has expanded into more specialized tasks such as spondylolisthesis detection, grading, and segmentation [6-8]. Yet an important gap remains. Much of the existing literature is driven by the pursuit of higher accuracy alone, as if the best model is simply the one with the highest number. In practice, that assumption is incomplete. A model that is accurate but computationally heavy, slow, or difficult to deploy may be of limited value in real clinical screening settings. Moreover, much prior work treats scoliosis and spondylolisthesis as separate problems, whereas real screening scenarios often require a unified decision across multiple classes.

This gap becomes even more important when lightweight models are considered. Compact CNN architectures are attractive because they reduce memory demand, model size, and inference cost, all of which matter in resource-constrained or high-throughput clinical

https://doi.org/10.31436/iiumej.v27i2.4341

environments. However, lightweight efficiency often comes with a representational cost: smaller networks may overlook subtle anatomical cues that distinguish a mildly abnormal spine from a normal one. Attention mechanisms provide a promising way to address this tension. Modules such as Squeeze-and-Excitation and the Convolutional Block Attention Module (CBAM) help networks emphasize the most informative channels and spatial regions, thereby strengthening feature discrimination without a prohibitive computational burden [9-11]. In other medical imaging tasks, including chest X-ray analysis and retinal disease classification, such mechanisms have improved performance with relatively modest overhead [12-14]. Even so, their role in lightweight spinal X-ray classification remains insufficiently understood, particularly when the evaluation extends beyond accuracy to include efficiency, robustness, and deployment relevance.

Against this background, this study asks a practical but underexplored question: in spinal X-ray classification, what kind of model is not only accurate but also efficient enough to matter? To answer this question, this paper proposes an efficiency-aware deep learning framework for three-class classification of spinal disorders from X-ray images. Five lightweight baseline architectures, namely EfficientNetB0, MobileNetV2, MobileNetV3Small, MNASNet, and DenseNet121, are systematically evaluated and then contrasted with three CBAM-augmented variants and a dual-branch fusion model. Rather than reporting performance solely by accuracy, the study adopts a broader evaluation lens that includes F1-score, parameter count, floating-point operations (FLOPs), model size, latency, and throughput. To improve the credibility of the findings, the experiments are conducted using stratified 5-fold cross-validation, multiple class-balancing strategies, and targeted ablation studies on attention design, fine-tuning depth, fusion strategy, and data balancing. This design allows the study to move beyond leaderboard-style comparison and instead uncover why certain models work better, under what conditions, and at what computational cost.

The contribution of this work extends beyond applying another deep learning model to spinal imaging. Its main contributions are as follows:

- *It reframes the central question from accuracy alone to practical value*. The study proposes an efficiency-aware transfer learning framework for unified three-class spinal disorder classification and compares nine models using both effectiveness and efficiency metrics, including F1-score, parameter count, FLOPs, model size, latency, and throughput, under stratified 5-fold cross-validation and multiple class-balancing strategies, judging models by how lightly and quickly they operate as well as how well they predict.
- *It shows that attention is not magic, but conditional*. By integrating CBAM into lightweight CNNs, this work demonstrates that attention mechanisms do not automatically improve performance in every setting; their benefit becomes convincing only when paired with sufficient data-driven regularization, especially augmentation-based training.
- *It introduces a lightweight fusion strategy that is both strong and purposeful*. The proposed Fusion_MNv3_MNAS model combines complementary feature representations from two efficient backbones to improve robustness and achieve the best overall classification performance, while still avoiding the heavy computational burden of larger architectures.
- *It explains why performance changes, not just which model wins*. Through targeted ablation studies on attention components, fine-tuning depth, fusion design, and balancing strategy, this work shows that performance in spinal X-ray classification depends less on model size alone than on the fit between architecture, regularization, and training strategy.

https://doi.org/10.31436/iiumej.v27i2.4341

2. RECENT ADVANCES IN SPINAL DISORDER CLASSIFICATION

Deep learning has changed the landscape of spinal X-ray analysis. What was once dominated by manual measurements and clinician-dependent interpretation is now increasingly supported by models that can learn diagnostic patterns directly from radiographs. This shift is important because disorders such as scoliosis and spondylolisthesis are not minor visual irregularities. They are clinically significant conditions that can alter posture, function, pain, and long-term quality of life if not recognized early. In recent years, the literature has expanded rapidly across classification, detection, and segmentation tasks for spinal abnormalities [1, 6, 15]. At the same time, lightweight architectures and attention mechanisms have gained momentum because they enable models that are not only accurate but also efficient enough for practical deployment. Even so, the literature remains fragmented. Some studies emphasize predictive accuracy almost exclusively, others focus on single diagnostic tasks, and still others adopt architectures whose deployment costs are rarely discussed. As a result, the field has progressed quickly, but not always coherently. This section, therefore, reviews prior work through three connected themes: deep learning for spinal X-ray classification, lightweight architectures and efficiency-aware evaluation, and attention mechanisms for feature enhancement. The aim is not merely to summarize earlier studies, but to show what the field has achieved, what it has overlooked, and why a more balanced framework is now needed.

2.1. Deep Learning for Spinal X-Ray Classification

The earliest breakthroughs in this area established an important fact: spinal disorder classification from X-ray images can be automated with remarkably high performance. A notable study demonstrated the feasibility of three-class classification among normal spine, scoliosis, and spondylolisthesis using transfer learning on a dataset of 338 subjects, reporting a maximum accuracy of 98.02% [1]. That work was influential not only because of its strong performance, but also because it recognized that accuracy alone can be misleading in imbalanced datasets and highlighted F1-score as a more reliable measure. Using the same benchmark dataset, subsequent work explored different architectures and optimization strategies and reported an accuracy of 99.01% with AlexNet using stochastic gradient descent with momentum [5]. Together, these studies showed that deep learning was not merely promising in spinal X-ray classification. It was already highly competitive. Yet they also raised a deeper question. When several models report near-perfect results on a relatively small dataset, the issue is no longer whether deep learning works, but how robust those results are, how much they depend on the training setup, and whether the gains remain meaningful beyond a single benchmark split.

The literature has also expanded beyond simple classification into detection and segmentation tasks. A Faster R-CNN framework with a ResNet backbone was introduced for automated scoliosis diagnosis, showing that region-based deep learning models could assist or potentially replace parts of manual radiographic assessment [15]. More recent studies pushed toward larger datasets and more specialized objectives. For instance, a YOLOv8-based segmentation model was developed for spondylolisthesis detection using over 10,000 images, achieving a precision of 96.77% [7]. Another study combined deep learning with geometric analysis for grading spondylolisthesis and reported accuracies above 94% on both internal and external datasets [8]. In parallel, lightweight architectures such as EfficientNet and MobileNet variants were explored in musculoskeletal imaging tasks, suggesting that compact models can remain highly competitive while reducing computational burden [2, 14, 16]. These developments point to a clear trend: spinal imaging research is moving toward automated systems that are faster, broader, and more clinically relevant.

https://doi.org/10.31436/iiumej.v27i2.4341

However, the dominant pattern in prior work remains accuracy-centric. Many studies report high performance, often using single train-test splits, without systematically examining computational efficiency, fold-to-fold variance, or deployment constraints [1, 5]. In addition, a large part of the literature remains task-specific, focusing on either scoliosis or spondylolisthesis in isolation rather than on unified multi-class screening. This matters because real clinical workflows do not always present a single isolated question. They often require a model to distinguish among several plausible conditions within one decision framework. The literature has therefore shown that automated spinal analysis is possible, but it has not yet fully established what makes such systems robust, efficient, and practical for real-world clinical use.

2.2. Lightweight Architectures and Efficiency-Aware Evaluation

As deep learning matured, a more practical concern came into view: a model that performs well in a paper may still be unsuitable in practice if it is too large, too slow, or too resource-intensive to deploy. This concern has driven the rise of lightweight convolutional neural network architectures that seek to preserve strong predictive performance while reducing memory demand and inference cost. EfficientNet introduced compound scaling to coordinate network depth, width, and input resolution more systematically [17]. MobileNetV2 reduced computational load through inverted residual blocks and linear bottlenecks [18]. MobileNetV3 extended this efficiency-driven philosophy by integrating squeeze-and-excitation modules and neural architecture search for mobile and embedded environments [19]. MNASNet pushed the idea further by explicitly optimizing latency on target platforms [20]. DenseNet121, although architecturally different, improved feature reuse and gradient flow through dense connectivity while still remaining relatively compact compared with heavier classical backbones [21]. Together, these models represent a broader shift in deep learning design, from raw capacity toward purposeful efficiency.

This shift is especially meaningful in medical imaging. Clinical systems are not judged solely by peak accuracy but also by their ability to operate reliably under practical constraints. Several studies on chest X-ray analysis, retinal disease detection, and histopathology have shown that lightweight CNNs can achieve performance comparable to larger models while requiring far less computation [22, 23]. In structured imaging domains, this finding is particularly striking because it challenges the common intuition that deeper and larger models are always superior. In some cases, compact models perform competitively precisely because the visual domain is constrained and clinically relevant patterns are more structured than those in natural image benchmarks [23]. EfficientNet-based models, for example, have shown favorable accuracy-efficiency trade-offs in clinical classification settings [24]. These findings suggest that model selection should not be driven solely by accuracy, but by the balance between predictive performance and operational feasibility.

Despite this growing recognition, efficiency-aware evaluation is still underused in spinal X-ray analysis. Many studies continue to report classification performance without accompanying measures such as parameter count, floating-point operations, model size, latency, or throughput. This omission is not trivial. In healthcare, a model that is accurate but difficult to scale or too slow for practical integration may have limited clinical value. Recent discussions in medical artificial intelligence increasingly emphasize that deployable systems must be not only accurate, but also interpretable, scalable, and computationally feasible [25, 26]. Within this context, efficiency-aware evaluation is no longer a secondary consideration. It is becoming part of a responsible model assessment.

https://doi.org/10.31436/iiumej.v27i2.4341

2.3. Attention Mechanisms in Medical Imaging

If lightweight models solve one problem, they can also create another. Their compactness often comes at the cost of reduced representational capacity, which may limit their ability to capture subtle diagnostic cues. In spinal radiographs, such cues can be clinically decisive. A mild scoliotic curvature or a slight vertebral displacement may be easy to miss if the model does not learn where to focus. Attention mechanisms have emerged as an effective approach to this problem because they enable networks to emphasize informative features without significantly increasing computational cost. One of the earliest influential designs was the Squeeze-and-Excitation Network, which introduced channel-wise attention by adaptively recalibrating feature responses [10]. The Convolutional Block Attention Module extended this idea by combining channel attention with spatial attention, enabling a model to learn both which features matter and where the most relevant image regions are located [9, 11]. This combination makes CBAM particularly attractive for lightweight CNNs, where efficiency remains essential.

The usefulness of attention mechanisms has already been demonstrated across a broad range of medical imaging tasks. Attention-guided CNNs have improved thoracic disease classification from chest X-rays by directing the model toward clinically relevant regions [12]. Attention modules have also enhanced performance and interpretability in general image classification settings [27]. Attention U-Net improved segmentation by emphasizing salient anatomical structures [13], while related attention-based designs in retinal imaging improved classification accuracy with relatively small computational overhead [28]. Across these tasks, the message is consistent: attention can make models more discriminative without substantially increasing their weight. This is especially valuable when model efficiency is a design constraint rather than an afterthought.

In addition to attention mechanisms, fusion and multi-branch architectures have been investigated to improve robustness by combining complementary feature representations. Multi-branch networks can capture different feature scales and contextual patterns, and their outputs can be aggregated through strategies such as feature concatenation or logit averaging [29, 30]. These designs are particularly appealing in medical imaging, where limited data and subtle class boundaries often mean that no single representation is sufficient. Yet despite the growing success of attention and fusion strategies in other imaging domains, their application to spinal X-ray classification remains limited. This gap is important because it suggests that the field has explored many useful ideas independently, but has not yet fully tested how they interact within a unified and efficiency-aware framework.

2.4. Research Gap Analysis and Study Positioning

The literature on spinal X-ray analysis has reached an interesting stage. The field has already shown that deep learning can work, sometimes with strikingly high reported performance. Early studies demonstrated that transfer learning could classify normal spine, scoliosis, and spondylolisthesis with very high accuracy on a dataset of 338 subjects [1, 5]. Other studies expanded the scope to include scoliosis detection, spondylolisthesis segmentation, and grading using larger datasets and more specialized pipelines [7, 8, 15]. At first glance, this progress suggests the problem is nearly solved. Yet a closer reading shows that the story is less complete. High accuracy has been reported, but often under narrow evaluation settings, with limited discussion of robustness, efficiency, or deployment cost. In other words, the field has produced strong answers to the question of whether spinal disorder classification is possible, but weaker answers to the more important question of which solutions are reliable and practical enough to matter in real screening settings.

Three gaps remain especially clear. First, much of the literature is still driven by an accuracy-centric evaluation culture. Reported results are often based on single train-test splits, and therefore say little about statistical stability or sensitivity to data partitioning [1, 5]. This becomes particularly important when the benchmark dataset is relatively small, because even strong headline results may reflect favorable splits rather than consistently robust behavior. Second, computational efficiency is rarely treated as a core outcome. Most prior studies emphasize predictive performance while omitting measures such as parameter count, floating-point operations, model size, and inference latency, even though these factors strongly influence whether a model can be deployed in real clinical environments. Third, prior work is frequently task-specific. Many studies address scoliosis alone, spondylolisthesis alone, or a specialized subtask, such as segmentation or grading [2, 7, 8, 15], whereas practical screening requires a unified framework that can distinguish among multiple clinically plausible classes.

Recent progress in lightweight architecture design and attention mechanisms points toward a more meaningful direction, but important combinations remain underexplored. Lightweight backbones such as MobileNet and EfficientNet suggest that strong performance does not always require heavy networks [2, 14, 16, 24]. Attention mechanisms such as SE-Net and CBAM have shown that feature selection can be made more discriminative without incurring prohibitive computational costs [9, 14]. Fusion strategies have also shown promise in other medical imaging tasks by combining complementary representations to improve robustness [29, 30]. However, these elements have seldom been studied together in spinal X-ray classification under a unified and efficiency-aware protocol. This distinction is important. The novelty of the present study does not lie in claiming that CBAM alone is new, or that fusion alone is unprecedented. Rather, its novelty lies in bringing together lightweight backbones, CBAM-based enhancement, dual-branch fusion, class-balancing strategies, stratified 5-fold cross-validation, and efficiency-aware evaluation into a single coherent framework, and then analyzing not only which model performs best but also why performance changes across architectures and training conditions.

Table 1 sharpens this point by comparing representative prior studies not only in terms of task and best result, but also in terms of evaluation setting, efficiency reporting, and practical limitations. The contrast is revealing. Prior studies have made important contributions, but they largely optimize one dimension at a time: accuracy, detection capability, geometric precision, or task specialization. What remains missing is a framework that jointly addresses unified three-class classification, robustness through cross-validation, and deployability through explicit efficiency analysis. This is precisely where the present study is positioned.

To address these gaps, this study proposes an efficiency-aware deep learning framework for three-class classification of spinal disorders from X-ray images. The framework integrates lightweight CNN backbones with CBAM-based attention and a dual-branch fusion strategy, and evaluates them using stratified 5-fold cross-validation, multiple class-balancing strategies, and targeted ablation studies. More importantly, it treats model performance as something to interpret rather than merely announce. The central claim is therefore stronger than a simple accuracy comparison: in spinal X-ray classification, performance is governed not only by architecture but also by the interaction among architecture, regularization, and evaluation design. By jointly addressing effectiveness, efficiency, and robustness, this study aims to move the field from isolated high-scoring models toward more credible, explainable, and clinically deployable solutions.

Table 1. Comparison of Representative Deep Learning Approaches for Spinal X-Ray Analysis

Study	Main Task	Dataset	Approach	Best Reported Result	Evaluation Setting	Efficiency Reported	Key Limitation
Fraiwan *et al*., 2022 [1]	Three-class classification	338 subjects	Transfer learning CNN	98.02% accuracy	Fixed split	No	No cross-validation; limited generalizability analysis; no efficiency evaluation
Güneş *et al*., 2024 [5]	Three-class classification	338 subjects	AlexNet + SGDM	99.01% accuracy	Fixed split	No	Single-split evaluation; accuracy-centric comparison; no efficiency metrics
Chen *et al*., 2022 [15]	Scoliosis detection/classification	Spine X-rays	Faster R-CNN + ResNet	Strong automated detection performance	Task-specific evaluation	No	Heavy architecture; high computational cost; limited real-time practicality
Vephasayanant *et al*., 2024 [7]	Spondylolisthesis detection/segmentation	10,616 images	YOLOv8 segmentation	96.77% precision	Task-specific evaluation	No	Segmentation task; not directly comparable to three-class classification; no efficiency analysis
Xu *et al*., 2025 [8]	Spondylolisthesis diagnosis/grading	4,400 images	Deep learning + geometric analysis	96.1% accuracy	Internal and external datasets	No	Multi-stage pipeline; increased complexity; limited deployment feasibility
Maaliw, 2023 [2]	Scoliosis severity/Cobb angle estimation	Spine X-rays	Deep learning regression (SCOLIONET)	High correlation with ground truth	Regression setting	No	Regression-only task; not a multi-class classifier; no efficiency evaluation
This study	Three-class classification	338 subjects	CBAM-enhanced lightweight CNNs + dual-branch fusion	Comprehensive effectiveness-efficiency evaluation	Stratified 5-fold cross-validation with ablation studies	Yes	Addresses unified classification, robustness, and deployment-oriented evaluation

3. RESEARCH METHODOLOGY

This study was designed to answer a practical question with scientific discipline: which deep learning strategy can classify spinal disorders accurately while remaining efficient enough for realistic deployment? To answer that question, a structured experimental framework was developed that moves from data preparation to evaluation in a transparent and reproducible manner. The overall pipeline consists of dataset preparation, image preprocessing, model

https://doi.org/10.31436/iiumej.v27i2.4341

development, training under multiple class-balancing strategies, and performance assessment using both effectiveness and efficiency metrics. Rather than treating methodology as a sequence of disconnected technical steps, the framework was intentionally designed to address specific challenges in spinal X-ray classification, namely, limited data, class imbalance, fairness in comparison, and deployment relevance.

3.1. Experimental Setup

All experiments were conducted on a high-performance computing platform to ensure consistency, reproducibility, and fair comparison across all evaluated models. The implementation was developed using the PyTorch deep learning framework, selected for three practical reasons. First, it offers strong native CUDA support for GPU-accelerated training on Windows platforms. Second, its eager execution mode facilitates transparent debugging and flexible model development. Third, it has become one of the most widely adopted frameworks in contemporary deep learning research, making the implementation easier to inspect, reproduce, and extend. The complete hardware and software configuration used in this study is summarized in Table 2.

Table 2. Hardware and Software Configuration

Component	Specification
Operating System	Windows Server Data Center 2022 (Version 21H2, Build 20348.4893)
System Architecture	64-bit
CPU	Intel Core i9-14900K (32 threads)
RAM	64 GB DDR5
Storage	2 TB SSD + 4 TB HDD
GPU	NVIDIA GeForce RTX 4090 (24 GB VRAM)
CUDA Version	12.4
Python	3.11.15
PyTorch	2.3.1
torchvision	0.18.1
NumPy	1.26.4
Environment	Conda (Scoliosis)
Reproducibility Seed	42
Deterministic Setting	`cudnn.deterministic = True`
Benchmark Setting	`cudnn.benchmark = False`

Reproducibility was treated as a methodological requirement rather than an afterthought. A fixed random seed of 42 was applied across Python, NumPy, and PyTorch, including both CPU and CUDA operations. In addition, deterministic behavior was enforced by setting `torch.backends.cudnn.deterministic = True` and disabling runtime performance benchmarking through `torch.backends.cudnn.benchmark = False`. These settings reduce non-deterministic variation during training and help ensure that experimental outcomes can be reproduced consistently across repeated runs. This is particularly important in a study comparing multiple lightweight architectures, where small performance differences should reflect model behavior rather than random execution.

To further strengthen transparency and reproducibility, the full source code of the proposed framework has been made publicly available at the project repository available at https://github.com/tsgunawan/SpinalDisordersClassification. The code is released under the MIT License, allowing unrestricted use, modification, and redistribution with minimal legal restriction. By making both the implementation and configuration explicit, this study aims to support not only result verification but also future extension by other researchers.

https://doi.org/10.31436/iiumej.v27i2.4341

3.2. Dataset

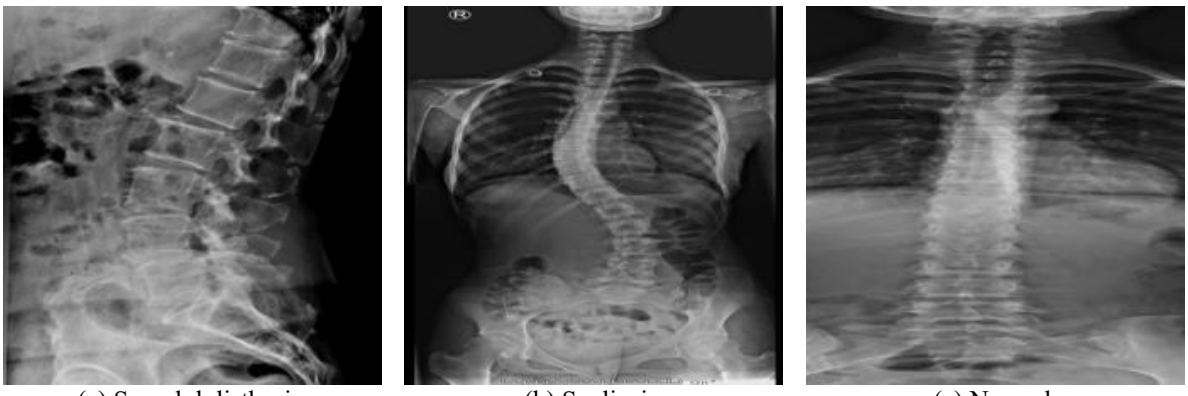
This study uses a publicly available spinal X-ray dataset introduced in [1], which has been widely adopted in prior work on automated classification of spinal disorders. The dataset contains 338 subjects and reflects a clinically meaningful screening scenario involving three diagnostic categories: Normal, Scoliosis, and Spondylolisthesis. Although the dataset is modest in size, this is precisely what makes it important. Small clinical datasets are common in real medical AI research, and they create a demanding setting in which model performance depends not only on architecture, but also on how carefully the experiment is designed. In this study, the limited dataset size was therefore treated not as a weakness to be ignored, but as a condition that required stricter evaluation and stronger methodological control.

To avoid inheriting bias from earlier evaluation protocols, the original dataset splits used in prior studies were not reused. Instead, all available samples were pooled and repartitioned using stratified 5-fold cross-validation. This strategy preserves the class proportion in each fold while making fuller use of the available data. Within each fold, the training portion was further divided into training and validation subsets using an 80:20 stratified split. This design provides a more reliable estimate of generalization performance than a single train-test split, especially when the dataset is small and imbalanced. It also allows performance to be reported as mean and standard deviation across folds, which gives a more credible picture of both central tendency and variability.

Table 3. Class Distribution of the Spine X-Ray Dataset

Class	**Subjects**	**Female**	**Male**	**Proportion**
Normal	71	40	31	21.0%
Scoliosis	188	151	37	55.6%
Spondylolisthesis	79	49	30	23.4%
Total	338	240	98	100%

As shown in Table 3, the dataset is inherently imbalanced, with scoliosis forming the majority class. This class distribution is clinically plausible, but it also introduces an important learning challenge. Without proper control, the model may become biased toward the dominant category and underperform on minority classes that are equally important from a diagnostic perspective. For this reason, multiple class-balancing strategies were incorporated into the experimental design, as described in Section 3.4, so that model performance could be examined under different responses to imbalance rather than under a single fixed assumption.



(a) Spondylolisthesis (b) Scoliosis (c) Normal

Figure 1. Sample Images from Each Class in the Spinal X-Ray Dataset

https://doi.org/10.31436/iiumej.v27i2.4341

Figure 1 presents representative X-ray images from the three diagnostic classes. These examples offer a useful visual reminder that the task is not trivial. The differences among classes are medically meaningful but, in some cases, visually subtle, especially when abnormality is mild or structural change is localized. The figure, therefore, illustrates why strong feature extraction and robust generalization are essential for deep learning-based spinal disorder classification. The challenge is not simply to detect obvious deformity, but to learn stable discriminative patterns across images that vary in anatomy, severity, and visual presentation.

3.3. Preprocessing and Augmentation

In a small medical imaging dataset, preprocessing is not a minor technical detail. It is the stage that determines whether models learn stable diagnostic patterns or become distracted by irrelevant variation. For that reason, this study adopts a unified preprocessing and augmentation pipeline across all evaluated models. The purpose is twofold: first, to ensure a fair comparison among architectures, and second, to fully leverage transfer learning from ImageNet-pretrained backbones. By standardizing the inputs before model-specific learning begins, the experiment reduces avoidable bias and keeps the comparison focused on what matters most, namely how each architecture responds to the same diagnostic problem.

3.3.1. Input Standardization

All input images were resized to a fixed resolution of 224 × 224 pixels to ensure compatibility with standard ImageNet-pretrained CNN architectures, including EfficientNet, MobileNet, and DenseNet variants [17, 19]. A common normalization scheme was then applied using the ImageNet mean and standard deviation values, namely mean = [0.485, 0.456, 0.406] and standard deviation = [0.229, 0.224, 0.225]. This choice aligns the input distribution with the pretrained model weights, thereby supporting more stable convergence during training [31]. More importantly, it eliminates inconsistencies that might otherwise arise from model-specific preprocessing choices. In other words, every model begins from the same visual footing, making the comparison more credible and methodologically cleaner.

3.3.2. Training Data Augmentation

Because the dataset is limited in size, the models must learn from clinically plausible variation rather than from repeated exposure to nearly identical images. To improve generalization and reduce overfitting, stochastic data augmentation was applied during training. The augmentation pipeline includes the following transformations:

- `RandomRotation` (±20°)
- `RandomResizedCrop` (scale = [0.85, 1.15])
- `ColorJitter` (brightness = 0.1, contrast = 0.1)
- `RandomHorizontalFlip`

These transformations introduce controlled changes in orientation, scale, and visual intensity while preserving the essential anatomical structure of the spine. The objective is not to aggressively distort the data, but to teach the model that clinically relevant patterns should remain recognizable even when the image presentation varies. This kind of augmentation has been widely shown to improve robustness in medical image classification by effectively increasing the diversity of training samples without collecting new data [32].

https://doi.org/10.31436/iiumej.v27i2.4341

3.3.3. Augmentation-Only Strategy

For the `augmentation_only` balancing strategy, the augmentation pipeline was expanded further to increase data diversity without relying on synthetic resampling. Two additional transformations were incorporated: `RandomAffine` and `RandomPerspective`. These operations simulate geometric perturbations and mild viewpoint variation, encouraging the model to develop more durable feature representations under realistic visual shifts. Unlike resampling-based techniques, this strategy works entirely with transformed versions of real images. As a result, it avoids the risk of generating artificial samples that may look mathematically plausible but anatomically unrealistic. This makes augmentation-only learning particularly attractive in medical imaging, where fidelity to real structure is critical.

3.3.4. Validation and Testing Protocol

For validation and testing, no augmentation was applied. All images were passed through a deterministic preprocessing pipeline that resized them to 224 × 224 pixels, converted them to tensor format, and normalized them using ImageNet statistics. This ensures that evaluation metrics reflect true model generalization rather than performance inflated by augmented inputs. The distinction is important: augmentation is used only to improve learning during training, not to make evaluation easier. In this way, the reported results remain a more faithful reflection of how the model behaves on unseen radiographs.

3.4. Class Balancing Strategies

Class imbalance is one of the most persistent challenges in medical image classification. When one category appears much more frequently than others, a model can achieve deceptively strong overall accuracy by learning the majority class well while performing poorly on clinically important minority classes. In a screening context, that imbalance is not a statistical inconvenience. It is a diagnostic risk. Formally, given a dataset $\mathcal{D} = [(x_i, y_i)]_{i=1}^{N}$ with class prior distribution $P(y)$, imbalance occurs when $P(y_k) \neq P(y_j)$ for classes $k \neq j$. Under such conditions, empirical risk minimization tends to bias the decision boundary toward the majority class, leading to poorer minority-class recognition and weaker generalization [33]. To address this issue, this study evaluates three complementary balancing strategies that act at different levels of the learning pipeline: loss-level adjustment, representation-level resampling, and input-level transformation. This design allows the study to examine not only whether balancing helps, but also which type of balancing works best for each model.

3.4.1. Class Weighting (class_weight)

The first strategy addresses the imbalance by modifying the loss function. Each class is assigned a weight inversely proportional to its frequency, so that minority classes contribute more strongly during optimization. Let C denote the number of classes, and w_c represent the weight assigned to the class c. The weight is defined as:

$$w_c = \frac{N}{C \cdot N_c} \tag{1}$$

where N is the total number of samples and N_c is the number of samples in class c. The weighted cross-entropy loss is then formulated as:

$$\mathcal{L}_{weighted} = -\sum_{i=1}^{N} \sum_{c=1}^{C} w_c y_{i,c} \log \hat{y}_{i,c} \tag{2}$$

where $y_{i,c}$ is the ground-truth indicator, and $\hat{y}_{i,c}$ is the predicted probability.

This strategy does not change the dataset itself. Instead, it changes the extent to which each class influences learning. Its main advantage is stability. It preserves the original data manifold,

https://doi.org/10.31436/iiumej.v27i2.4341

avoids synthetic sample generation, and provides a transparent baseline for imbalanced learning [33]. For that reason, class weighting is used in this study as the reference strategy against which the other balancing approaches are compared.

3.4.2. Feature-Space SMOTE (smote_features)

The second strategy extends the Synthetic Minority Over-sampling Technique (SMOTE) from the image domain into the learned feature space. In classical SMOTE, new minority samples are generated by interpolating between neighboring examples:

$$x_{\text{new}} = x_i + \lambda(x_j - x_i), \quad \lambda \sim \mathcal{U}(0,1) \tag{3}$$

where x_i and x_j are minority-class samples. While this is often useful for tabular data, direct interpolation in pixel space is problematic for medical X-ray images because it can produce anatomically implausible structures.

To avoid that problem, SMOTE is applied here not to raw images, but to deep feature embeddings $z = f_\theta(x) \in \mathbb{R}^d$, where f_θ denotes a pretrained CNN backbone and $d \in [576, 1280]$ depending on the architecture. Synthetic embeddings are generated as

$$z_{new} = z_i + \lambda(z_j - z_i) \tag{4}$$

This creates a two-stage learning process:

- Learn a feature extractor f_θ from input images
- Apply SMOTE in the feature space to obtain balanced embeddings
- Train a classifier $g_\phi(z)$ on the augmented feature set

By operating in a semantically structured latent space, feature-space SMOTE is more likely to preserve meaningful class relationships while improving minority representation [34]. However, this benefit comes with a trade-off. The training pipeline becomes decoupled, and the method is not directly compatible with end-to-end architectures that rely on attention modules or fusion mechanisms. This limitation is important and is made explicit in this study.

3.4.3. Augmentation-Only Strategy (augmentation_only)

The third strategy addresses class imbalance solely through stochastic augmentation, without explicit resampling or loss reweighting. Let $\mathcal{T}$ denote a distribution over valid image transformations such as rotation, scaling, or intensity adjustment. Augmented samples are produced as

$$x' = T(x), \quad T \sim \mathcal{T} \tag{5}$$

where the transformation preserves the semantic class label y.

This strategy enlarges the effective support of the input distribution and encourages the model to learn more invariant representations. In this study, geometric and photometric transformations, including affine transformations and perspective distortions, are used to enrich the training set. Unlike SMOTE-based methods, augmentation-only learning works directly with physically plausible image variations and therefore maintains anatomical realism. It also preserves end-to-end training, making it compatible with all baseline, attention-enhanced, and fusion-based architectures. This is one reason why augmentation-based learning is often especially valuable in small medical imaging datasets [32].

https://doi.org/10.31436/iiumej.v27i2.4341

3.4.4. Strategy Comparison and Rationale

The three balancing strategies explored in this study reflect three different philosophies of correction:

- **Class weighting** changes the optimization objective.
- **Feature-space SMOTE** augments the learned representation space.
- **Augmentation-only** enriches the input distribution.

A deliberate design decision in this work is the exclusion of pixel-space SMOTE. Although straightforward in principle, pixel-level interpolation can generate anatomically inconsistent X-ray images, potentially distorting the very structures the model is expected to learn. In contrast, feature-space SMOTE operates on semantically meaningful embeddings, while augmentation-based learning preserves physical plausibility through valid transformations of real images. This distinction is not only technical, but methodological. In medical imaging, a balancing strategy must not merely increase minority representation. It must do so without breaking the anatomical logic of the data. The downstream consequences of these three strategies, and their interaction with model architecture, are examined further through the ablation studies presented in Section 5.

3.5. Baseline Models

A fair benchmark should not compare models that all think the same way. It should compare models that represent genuinely different design philosophies, especially when the goal is to understand not only which model performs best, but why. For that reason, five convolutional neural network architectures were selected as baseline models: EfficientNetB0, MobileNetV2, MobileNetV3Small, MNASNet, and DenseNet121. These models span a meaningful range of modern design strategies, including compound scaling, inverted residual learning, hardware-aware search, platform-aware latency optimization, and dense feature reuse. Together, they create a spectrum from compact mobile-oriented networks to deeper feature-reuse architectures, making them well-suited for studying the trade-off between diagnostic performance and computational efficiency in spinal X-ray classification.

This selection was also made with transparency in mind. During the early model consideration, NAS-inspired lightweight networks were of particular interest due to their relevance to efficient deployment. In the final benchmark, MNASNet was used as the representative NAS-based architecture. This choice was made deliberately because MNASNet provides stable pretrained support within the PyTorch and torchvision environment used in this study, enabling a consistent and reproducible implementation across experiments. Thus, MNASNet serves here as the practical latency-aware NAS-based baseline for comparison. This clarification is important because benchmark credibility depends not only on what models are selected, but also on whether those selections are technically reproducible and openly justified.

Table 4 summarizes the selected architectures in terms of embedding dimension, approximate parameter count, ImageNet Top-1 accuracy, and design philosophy. EfficientNetB0 uses compound scaling to jointly balance network depth, width, and image resolution, achieving strong predictive performance with relatively modest complexity [17]. MobileNetV2 reduces computational burden through inverted residual blocks and linear bottlenecks, making it one of the most influential lightweight backbones in efficient vision modeling [18]. MobileNetV3Small further advances this idea by combining hardware-aware neural architecture search with squeeze-and-excitation modules, producing a model optimized for low-resource environments [19]. MNASNet extends the neural architecture search

https://doi.org/10.31436/iiumej.v27i2.4341

paradigm by directly targeting inference latency on deployment hardware [20]. DenseNet121, in contrast, follows a different logic: each layer receives information from all previous layers, encouraging feature reuse and smoother gradient flow while maintaining reasonable parameter efficiency [21]. These five backbones, therefore, do more than provide variety. They allow the benchmark to test whether spinal X-ray classification benefits more from compact efficiency, richer connectivity, or a carefully balanced compromise between the two.

Table 4. Baseline CNN Architectures and Their Characteristics

Model	Torchvision Function	Embedding Dimension	Parameters (Millions)	ImageNet Top-1 Accuracy (%)	Design Philosophy
EfficientNetB0	efficientnet_b0	1280	~5.3	77.1	Compound scaling of depth, width, and resolution
MobileNetV2	mobilenet_v2	1280	~3.4	71.9	Inverted residuals with linear bottlenecks
MobileNetV3Small	mobilenet_v3_small	576	~2.5	67.7	Hardware-aware NAS with squeeze-and-excitation
MNASNet	mnasnet1_0	1280	~4.4	73.5	Platform-aware neural architecture search
DenseNet121	densenet121	1024	~8.0	74.4	Dense connectivity with feature reuse

To ensure that backbone differences, rather than classifier design, drive the comparison, all baseline models were adapted using the same transfer learning framework. Let $f_\theta(x)$ denote the pretrained backbone that maps an input image x to a feature vector $z \in \mathbb{R}^d$, where d is the embedding dimension of the selected model. A lightweight classifier head $g_\phi(z)$ is then appended for three-class classification:

$$g_\phi(z) = W_2\,\sigma\big(Dropout(W_1 z)\big) \tag{6}$$

where $W_1 \in \mathbb{R}^{128\times d}$, $W_2 \in \mathbb{R}^{3\times 128}$, and $\sigma(\cdot)$ denotes the ReLU activation function. This corresponds to the following architecture:

- `Linear`$(d \rightarrow 128)$
- `ReLU`
- `Dropout(0.3)`
- `Linear`$(128 \rightarrow 3)$

This unified classifier head serves an important methodological purpose. It removes unnecessary variation at the output stage and ensures that performance differences can be interpreted primarily as consequences of backbone behavior rather than custom head engineering. In a benchmark study, this kind of control is essential because it keeps the comparison honest.

https://doi.org/10.31436/iiumej.v27i2.4341

The training strategy follows a two-stage transfer learning paradigm. In the first stage, all backbone parameters θ are frozen, and only the classifier head parameters ϕ are optimized. This allows each model to retain the general visual representations learned from ImageNet while adapting its final decision layer for spinal X-ray classification, reducing the risk of overfitting. In the second stage, described in Section 3.7, a subset of higher-level backbone layers is unfrozen to permit domain-specific refinement. This staged approach is especially important for small medical datasets, where training a deep network end-to-end from scratch is rarely practical and often unstable [23]. In effect, the model is first taught to speak the language of the dataset, and only then allowed to refine how it listens to the anatomy.

3.6. Proposed Models

The proposed models were designed around a simple but important idea: in spinal X-ray classification, a model should not become heavier merely to become stronger. Instead, it should learn to focus better. Guided by this principle, this study develops an efficiency-aware, attention-enhanced deep learning framework that combines lightweight convolutional backbones, the Convolutional Block Attention Module (CBAM), and a dual-branch fusion strategy for three-class classification of spinal disorders. As illustrated in Figure 2 and formalized in Algorithm 1, the framework aims to improve discriminative feature learning while preserving the computational efficiency needed for realistic deployment. The central design choice is therefore not brute-force depth, but selective enhancement. Rather than adding complexity indiscriminately, the model is encouraged to attend to the most informative spinal features and, when beneficial, to combine complementary viewpoints from two lightweight backbones.

3.6.1. CBAM-Augmented Backbones

The first component of the proposed framework strengthens conventional transfer learning models by inserting an attention module between the pretrained backbone and the classifier head. The overall pipeline follows a simple sequence:

$$\text{Input Image} \rightarrow \text{Frozen Backbone} \rightarrow \text{CBAM} \rightarrow \text{GAP} \rightarrow \text{Classifier Head} \quad (7)$$

Given an input X-ray image $x \in \mathbb{R}^{3\times224\times224}$, a pretrained backbone network f_θ extracts a feature map $F \in \mathbb{R}^{C\times H\times W}$. Because the dataset is limited, the backbone is initially frozen to preserve the general visual knowledge learned from ImageNet and to reduce overfitting during early training. The extracted features are then refined by CBAM, which applies channel attention followed by spatial attention. This ordering matters. The model is first encouraged to decide **which** features are important, and only then to determine **where** the relevant evidence lies in the image. In spinal radiographs, this is particularly valuable because pathological evidence is often subtle, localized, and easily diluted within a larger anatomical field.

The *channel attention module* aggregates global context using both average pooling and max pooling, followed by a shared multilayer perceptron and sigmoid activation:

$$M_c = \sigma\left(MLP\big(AvgPool(F)\big) + MLP\big(MaxPool(F)\big)\right) \quad (8)$$

The resulting channel-refined feature map is computed as $F_c = M_c \odot F$, where $\odot$ denotes element-wise multiplication. This mechanism helps the network amplify diagnostically informative channels while suppressing less relevant responses.

Next, the *spatial attention module* identifies the image regions that deserve emphasis. It applies channel-wise pooling and then a convolution with a 7×7 kernel:

https://doi.org/10.31436/iiumej.v27i2.4341

$$M_s = \sigma\left(\text{Conv}_{7\times7}([\text{AvgPool}(F_c); \text{MaxPool}(F_c)])\right) \quad (9)$$

The final refined feature map is then obtained as $F' = M_s \odot F_c$. This sequential attention process allows the model to emphasize both globally important features and spatially localized patterns, such as abnormal spinal curvature, vertebral misalignment, or subtle radiographic distortions associated with pathology.

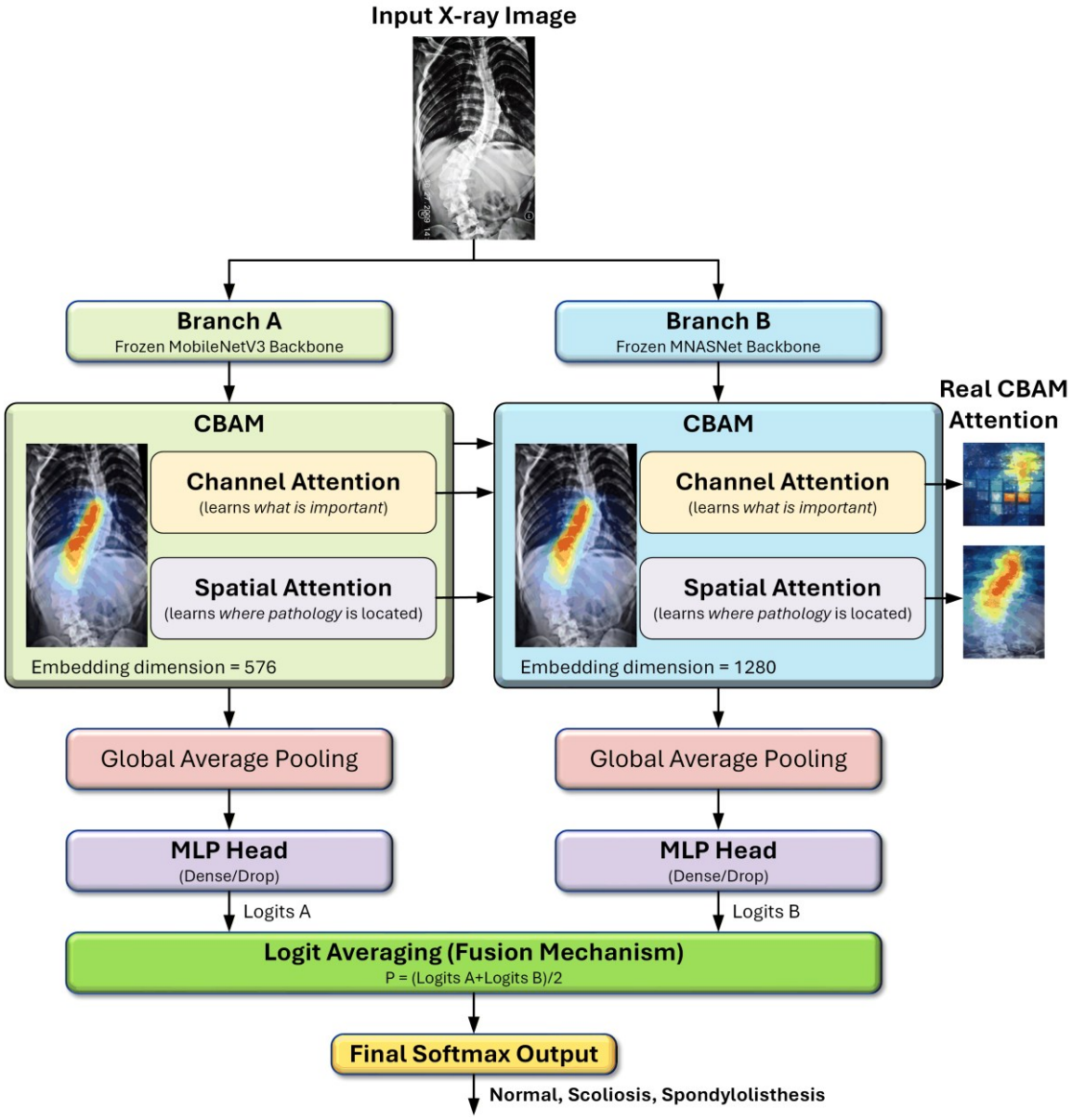


Figure 2. Proposed CBAM-Enhanced Dual-Branch Fusion Model for Multi-Class Spinal X-Ray Classification

https://doi.org/10.31436/iiumej.v27i2.4341

Algorithm 1. Inference Procedure of the Proposed CBAM-Based Dual-Branch Fusion Framework

Algorithm 1 Forward Pass of the Proposed Dual-Branch CBAM-Enhanced Fusion Model

Require: Input spine X-ray image $x \in \mathbb{R}^{3\times224\times224}$
Require: Frozen pretrained backbones f_A and f_B
Require: CBAM modules CBAM_A and CBAM_B
Require: Classifier heads g_A and g_B
Ensure: Predicted probability vector $\hat{\mathbf{y}} \in \mathbb{R}^3$ for {Normal, Scoliosis, Spondylolisthesis}

Branch A: CBAM-MobileNetV3Small
$F_A \leftarrow f_A(x)$ ▷ $F_A \in \mathbb{R}^{C_A\times H_A\times W_A}$
$F'_A \leftarrow \text{CBAM}(F_A)$ ▷ Attention-refined feature map
$\mathbf{z}_A \leftarrow \text{GAP}(F'_A)$ ▷ $\mathbf{z}_A \in \mathbb{R}^{576}$
$\mathbf{l}_A \leftarrow g_A(\mathbf{z}_A)$ ▷ $\mathbf{l}_A \in \mathbb{R}^3$

Branch B: CBAM-MNASNet
$F_B \leftarrow f_B(x)$ ▷ $F_B \in \mathbb{R}^{C_B\times H_B\times W_B}$
$F'_B \leftarrow \text{CBAM}(F_B)$ ▷ Attention-refined feature map
$\mathbf{z}_B \leftarrow \text{GAP}(F'_B)$ ▷ $\mathbf{z}_B \in \mathbb{R}^{1280}$
$\mathbf{l}_B \leftarrow g_B(\mathbf{z}_B)$ ▷ $\mathbf{l}_B \in \mathbb{R}^3$

Fusion by Logit Averaging
$\mathbf{l}_{\text{fuse}} \leftarrow 0.5 \cdot \mathbf{l}_A + 0.5 \cdot \mathbf{l}_B$

Final Prediction
$\hat{\mathbf{y}} \leftarrow \text{Softmax}(\mathbf{l}_{\text{fuse}})$
return $\hat{\mathbf{y}}$

function $\text{CBAM}(F)$
  $M_c \leftarrow \sigma\Big(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))\Big)$ ▷ Channel attention map
  $F_c \leftarrow M_c \odot F$ ▷ Channel-refined feature map
  $M_s \leftarrow \sigma\Big(\text{Conv}_{7\times7}([\text{AvgPool}(F_c); \text{MaxPool}(F_c)])\Big)$ ▷ Spatial attention map
  $F_s \leftarrow M_s \odot F_c$ ▷ Final refined feature map
  return F_s
end function

function $g(\mathbf{z})$
  $\mathbf{h} \leftarrow \text{ReLU}(W_1\mathbf{z} + \mathbf{b}_1)$
  $\tilde{\mathbf{h}} \leftarrow \text{Dropout}(\mathbf{h}, p = 0.3)$
  $\mathbf{l} \leftarrow W_2\tilde{\mathbf{h}} + \mathbf{b}_2$
  return $\mathbf{l}$
end function

Notation: f_A denotes the frozen MobileNetV3Small backbone and f_B denotes the frozen MNASNet backbone. $GAP(\cdot)$ is global average pooling, $\sigma(\cdot)$ is the sigmoid activation, and $\odot$ denotes element-wise multiplication. The classifier heads g_A and g_B share the same multilayer perceptron structure: Linear($d \rightarrow 128$) → ReLU → Dropout(0.3) → Linear($128 \rightarrow 3$).

After attention refinement, global average pooling converts the feature map into a compact vector $z \in \mathbb{R}^C$, where $C \in [576, 1024, 1280]$ depends on the backbone. A lightweight classifier head then produces the class logits. To keep the comparison controlled and fair, the same head design is used across all attention-enhanced models. In this study, CBAM is integrated into three backbones: MobileNetV3Small, MNASNet, and EfficientNetB0. These models were selected because they offer different trade-offs between compactness and

representational strength. The goal is not simply to test whether CBAM improves performance, but to examine under which architectural conditions attention becomes most useful.

3.6.2. Dual-Branch Fusion Model

Attention helps a model focus. Fusion helps it hesitate less. To further improve robustness and generalization, this study proposes a dual-branch fusion architecture named **Fusion_MNv3_MNAS**. As shown in Figure 2 and described in Algorithm 1, the same input image is processed through two parallel CBAM-enhanced branches:

- **Branch A**: CBAM-MobileNetV3Small, which provides a compact and efficient representation with $d = 576$.
- **Branch B**: CBAM-MNASNet, which provides a richer latency-aware representation with, $d = 1280$.

Each branch independently produces a logit vector:

$$\mathbf{l}_A = f_1(x), \mathbf{l}_B = f_2(x), \mathbf{l}_A, \mathbf{l}_B \in \mathbb{R}^3 \tag{10}$$

The final prediction is obtained through *logit-level fusion*:

$$\hat{\mathbf{y}} = \text{Softmax}\left(\frac{1}{2}\mathbf{l}_A + \frac{1}{2}\mathbf{l}_B\right) \tag{11}$$

This design is motivated by a simple insight from ensemble learning: when two competent learners make different kinds of mistakes, combining them can reduce variance and improve reliability. Here, the two branches are intentionally lightweight but not redundant. MobileNetV3Small favors compact, aggressively efficient feature extraction, whereas MNASNet contributes a somewhat richer representation shaped by latency-aware neural architecture search. Their diversity is therefore functional rather than cosmetic. The fusion model is not merely larger than either branch alone; it is more balanced because it merges two different perspectives on the same anatomy.

This distinction is important for the framework's novelty. The contribution is not the isolated use of CBAM or fusion, both of which have appeared in other domains. Rather, the contribution lies in combining lightweight backbones, attention refinement, and dual-branch fusion within a unified spinal X-ray classification framework that is explicitly evaluated under efficiency-aware criteria. In other words, the architecture is designed not just to score well, but to reveal how complementary lightweight models can cooperate without drifting into the cost profile of a heavy network.

From an efficiency standpoint, the proposed fusion model maintains a favorable parameter-performance balance and remains lighter than heavier baselines such as DenseNet121. Unlike conventional ensembles that require separately trained models and duplicated inference workflows, the proposed architecture integrates both branches into a single unified framework, allowing efficient parallel feature extraction and coordinated prediction. By embedding CBAM within each backbone, the model becomes more selective about the spinal patterns it amplifies. By combining two heterogeneous lightweight branches, it becomes more robust to representation bias and less dependent on a single architectural viewpoint. The resulting design, therefore, aims to achieve what is often difficult in medical AI: improve discrimination, strengthen robustness, and preserve practical deployment.

3.7. Training Protocol

Training on a small medical imaging dataset requires restraint as much as ambition. If the model is adapted too little, it may remain tied to generic ImageNet features that are not

sufficiently sensitive to spinal anatomy. If it is adapted too aggressively, it may overfit the limited training data and lose the stability gained from pretraining. To navigate this tension, this study adopts a two-phase training protocol that combines frozen feature extraction with selective fine-tuning. The strategy is designed to preserve the broad visual knowledge already learned by the pretrained backbone while gradually allowing the model to specialize in spinal X-ray interpretation. In this sense, the model is not forced to relearn vision from the beginning. It is first grounded, then refined.

3.7.1. Phase 1: Frozen Backbone Training

In the first phase, all backbone parameters θ were frozen, and only the classifier head parameters ϕ were optimized. This allows the model to immediately leverage the generic visual representations learned from large-scale ImageNet pretraining while minimizing the risk of overfitting to the relatively small spinal X-ray dataset. The model was trained for 50 epochs using the Adam optimizer with an initial learning rate of 1×10^{-4}. The optimization objective can be expressed as:

$$\phi^* = \arg\min_{\phi} \mathcal{L}\big(g_\phi\big(f_\theta(x)\big), y\big), \quad \theta \; fixed \tag{12}$$

This phase has a practical purpose: it teaches the classifier head to interpret the backbone features before the backbone itself is allowed to adapt. For small medical datasets, this is often the safest starting point because it prevents the network from immediately drifting toward unstable domain-specific fitting.

3.7.2. Phase 2: Partial Fine-Tuning

After initial convergence, a second training phase was introduced to adapt the higher-level feature representations to the target domain. In the main benchmark protocol, the top 20% of backbone layers were unfrozen, while the lower-level layers remained fixed to preserve generic feature extraction capabilities. The model was then trained for an additional 30 epochs using a reduced learning rate of 1×10^{-5}, allowing gradual and controlled refinement of domain-relevant features:

$$(\theta', \phi') = \arg\min_{\theta', \phi} \mathcal{L}\big(g_\phi\big(f_{\theta'}(x)\big), y\big) \tag{13}$$

where $\theta' \subset \theta$ denotes the subset of unfrozen parameters.

This staged optimization helps reduce catastrophic forgetting while enabling the network to capture finer anatomical cues that are specific to spinal radiographs. It is especially important in this study because the abnormalities of interest, such as mild curvature or vertebral displacement, may be too subtle to be fully represented by generic pretrained features alone.

For transparency, the 20% unfreezing fraction was used as the **default benchmark setting** across models to maintain a uniform and fair experimental protocol during the main comparative evaluation. A later ablation study in Section 5.2 showed that, for some models, especially CBAM-MobileNetV3Small, unfreezing approximately 30% of the backbone yields slightly better adaptation. This does not invalidate the main benchmark. Rather, it shows that the globally fixed protocol used for fair comparison is not always identical to the model-specific optimum discovered through later analysis. Accordingly, the benchmark results should be interpreted as results under a controlled common protocol, while the ablation study identifies the more refined task-specific optimum. This clarification directly addresses the potential ambiguity between the benchmark default and the later fine-tuning analysis.

3.7.3. Regularization and Training Control

To improve generalization and reduce the risk of overfitting, three training control mechanisms were incorporated:

- *Early stopping* with a patience of 10 epochs, monitored on validation loss, to terminate training when no further improvement is observed.
- *Adaptive learning rate scheduling* using `ReduceLROnPlateau`, which reduces the learning rate by a factor of 0.5 when validation loss fails to improve for 5 consecutive epochs.
- *Model checkpointing*, in which the model state achieving the best validation loss is saved to preserve the strongest generalizing configuration.

These mechanisms were included not simply to stabilize optimization, but to ensure that performance differences among models reflect learning quality rather than accidental overtraining. In a benchmark study involving multiple lightweight architectures, such control is essential for fair comparison.

3.8. Cross-Validation Strategy

A small dataset can produce impressively high numbers for the wrong reasons. One favorable train-test split may overestimate a model's true performance, while another may unfairly underestimate it. To avoid that problem, this study adopts a *stratified 5-fold cross-validation protocol* over the full dataset of 338 subjects. The purpose is not only to increase statistical reliability, but also to ensure that the reported results reflect repeated and balanced evaluation rather than a single lucky partition. Stratification preserves the original class distribution within each fold, so that all three diagnostic categories remain proportionally represented during both training and evaluation.

Let the dataset be denoted as $\mathcal{D} = [(x_i, y_i)]_{i=1}^{N}$, where $N = 338$. The data are partitioned into $K = 5$ mutually exclusive folds $[\mathcal{D}]_{k=1}^{5}$, such that:

$$\mathcal{D} = \bigcup_{k=1}^{5} \mathcal{D}_k \, , \mathcal{D}_i \cap \mathcal{D}_j = \emptyset \text{ for } i \neq j \tag{14}$$

At each iteration k, one fold $\mathcal{D}_k$ is reserved for evaluation, while the remaining four folds are used for model development:

$$\mathcal{D}_{\text{train}}^{(k)} = \mathcal{D} \setminus \mathcal{D}_k, \quad \mathcal{D}_{\text{val}}^{(k)} = \mathcal{D}_k \tag{15}$$

Within each iteration, the development portion is further internally stratified into *80% training and 20% validation* subsets. This additional split supports model selection, learning-rate adjustment, and early stopping without leaking information from the held-out evaluation fold. As a result, the evaluation remains separated from optimization, which is crucial for preserving the credibility of the final performance estimates.

Performance is aggregated across all folds using the mean and standard deviation:

$$\bar{M} = \frac{1}{K} \sum_{k=1}^{K} M^{(k)} \, , \sigma_M = \sqrt{\frac{1}{K} \sum_{k=1}^{K} \left(M^{(k)} - \bar{M} \right)^2} \tag{16}$$

where $M^{(k)}$ denotes the evaluation metric for fold k, such as accuracy or F1-score. Reporting results as $\bar{M} \pm \sigma_M$ serves two purposes at once. The mean captures the central performance level, while the standard deviation reveals how stable that performance remains across data

https://doi.org/10.31436/iiumej.v27i2.4341

partitions. In a dataset of this size, both are necessary. A high mean without stability can be misleading, whereas a slightly lower mean with low variability may represent the more trustworthy model. This is why cross-validation in this study is not merely a procedural choice. It is part of the argument for credibility.

3.9. Evaluation Metrics

A model that is accurate but impractical is difficult to trust in a real clinical pipeline. Conversely, a model that is fast but diagnostically weak is of little use to clinicians. For that reason, this study adopts a dual evaluation framework that treats performance from two angles at once: *effectiveness*, which reflects predictive quality, and *efficiency*, which reflects computational cost. This distinction is central to the present work. The objective is not merely to identify the model with the highest score, but to determine which model offers the most credible balance between diagnostic capability and deployment feasibility. In this sense, evaluation is not the last step of the framework. It is part of the paper's argument.

3.9.1. Effectiveness Metrics

Model effectiveness is evaluated using standard classification metrics derived from the confusion matrix. For each class c, let TP_c, FP_c, FN_c, and TN_c denote the numbers of true positives, false positives, false negatives, and true negatives, respectively.

The overall accuracy is defined as:

$$\text{Accuracy} = \frac{\sum_c TP_c}{\sum_c (TP_c + FP_c + FN_c + TN_c)} \tag{17}$$

Accuracy provides a useful global summary of correctness, but on an imbalanced dataset, it can also be misleading. A model may achieve a high overall score simply by favoring the majority class, even while underperforming on clinically important minority classes. For that reason, this study gives particular importance to *weighted metrics* that account for class distribution and therefore offer a more faithful view of model behavior across all diagnostic categories.

The class-wise *precision*, *recall*, and *F1-score* are defined as:

$$\text{Precision}_c = \frac{TP_c}{TP_c + FP_c}, \quad \text{Recall}_c = \frac{TP_c}{TP_c + FN_c}, \quad \text{F1}_c = \frac{2 \times \text{Precision}_c \times \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c} \tag{18}$$

The *weighted averages* are computed as:

$$\text{Metric}_{\text{weighted}} = \sum_{c=1}^{C} w_c \cdot \text{Metric}_c, w_c = \frac{N_c}{N} \tag{19}$$

where N_c is the number of samples in class c, and N is the total number of samples. This weighting ensures that each class contributes in proportion to its presence in the dataset, thereby reducing bias toward the majority class.

In this study, weighted precision, weighted recall, weighted F1-score, and accuracy are reported consistently across folds, with **means ± standard deviations**. This consistency is important because a metric without variability can appear more certain than it truly is. Reporting dispersion across folds allows the reader to judge not only how well a model performs on average but also how stable that performance is across different data partitions.

To provide a more concrete view of model behavior, confusion matrices are also reported. They reveal where the model succeeds, where it struggles, and which classes are most commonly confused. This is particularly useful in spinal X-ray classification, where mild scoliosis, borderline normal cases, or subtle vertebral displacement may create ambiguity that

summary scores alone cannot fully explain. In addition, *Receiver Operating Characteristic* (ROC) curves and the corresponding *Area Under the Curve* (AUC) are computed using a one-vs-rest strategy. The ROC curve traces the true positive rate against the false positive rate across decision thresholds, while the AUC summarizes the model's class-separation ability independent of any single threshold. Together, these metrics provide a fuller picture of diagnostic discrimination.

3.9.2. Efficiency Metrics

Predictive performance alone does not determine whether a model is useful in practice. In clinical deployment, a model must also be compact enough to store, fast enough to respond, and efficient enough to run within realistic hardware constraints. For this reason, computational efficiency is evaluated systematically alongside classification quality.

Three structural efficiency metrics are considered:

- *Parameter Count*, including both total and trainable parameters, to quantify model complexity and memory demand.
- *Floating Point Operations (FLOPs)*, computed using the `thop` library, to estimate the computational burden of inference.
- *Model Size*, measured in megabytes (MB) based on the serialized `state_dict`, to reflect storage requirements.

In addition to structural cost, runtime behavior is evaluated using two complementary inference metrics:

- *Latency*, defined as the time required to process a single input sample. Latency is computed as the mean inference time over T repeated runs after an initial warm-up phase:

$$\text{Latency} = \frac{1}{T}\sum_{t=1}^{T} t_{\text{inference}}^{(t)} \tag{20}$$

- *Throughput*, defined as the number of images processed per second:

$$\text{Throughput} = \frac{\text{Number of images}}{\text{Total inference time}} \tag{21}$$

To improve measurement stability, latency is recorded over 50 runs with 5 warm-up iterations to reduce transient GPU effects. All runtime measurements are performed under the same hardware conditions and with batch size 1, so that the evaluation reflects a realistic clinical screening scenario in which patient scans are processed individually rather than in large offline batches.

This efficiency analysis is not included as a secondary supplement. It is integral to the study's central claim. A model with marginally higher accuracy may not represent the better solution if it requires substantially more parameters, more memory, and more inference time. By reporting both predictive effectiveness and computational efficiency side by side, this study aims to identify models that are not only theoretically strong but also credible for real-world deployment in spinal disorder screening.

4. RESULTS AND DISCUSSION

The value of a medical AI model is revealed not only by the score it reaches, but by the trade-offs it makes to get there. For that reason, this section evaluates the proposed framework from two connected perspectives: effectiveness and efficiency. All results are reported using

stratified 5-fold cross-validation as the mean ± standard deviation, so that performance is judged not only by its absolute value but also by its stability across data partitions. This distinction matters for a dataset of 338 subjects, where a single favorable split can make a model appear more certain than it truly is. The discussion, therefore, moves beyond reporting which model scored highest. It examines which models remain credible under repeated evaluation, which architectures use computation most wisely, and where the remaining classification errors still concentrate.

4.1. Baseline Comparison

Table 5 compares the five baseline architectures under the `class_weight strategy`. The first important finding is that the top three baselines are essentially tied in predictive performance. MobileNetV3Small achieves a weighted F1-score of 0.962±0.022, while MobileNetV2 and EfficientNetB0 each achieve 0.961 with comparable variability. These differences are too small to support any strong claim of superiority based solely on accuracy. In practical terms, the baseline comparison does not identify one clearly dominant scorer. Instead, it shifts the question to something more meaningful: when predictive performance is nearly the same, which model earns that performance with the least computational burden?

Table 5. Performance and Efficiency Comparison of Baseline Models under Class-Weighted Strategy (5-fold stratified cross-validation, mean ± standard deviation)

Model	Accuracy	Precision (w)	Recall (w)	F1 (w)	Params (M)	FLOPs (M)	Size (MB)	Latency (ms)	Throughput (img/s)
EfficientNetB0	0.962 ± 0.022	0.962 ± 0.023	0.962 ± 0.022	0.961 ± 0.023	4.172	414.029	15.915	18.73	53.391
MobileNetV2	0.962 ± 0.027	0.962 ± 0.027	0.962 ± 0.027	0.961 ± 0.027	2.388	326.371	9.11	12.16	82.236
MobileNetV3Small	**0.962 ± 0.022**	**0.963 ± 0.022**	**0.962 ± 0.022**	**0.962 ± 0.022**	1.001	60.938	3.819	15.916	62.831
MNASNet	0.902 ± 0.030	0.923 ± 0.021	0.902 ± 0.030	0.905 ± 0.029	3.267	335.127	12.461	11.472	87.169
DenseNet121	0.935 ± 0.037	0.941 ± 0.034	0.935 ± 0.037	0.936 ± 0.036	7.085	2896.115	27.029	39.426	25.364

From that perspective, MobileNetV3Small stands out as the most persuasive baseline. It reaches the highest weighted F1-score while using only 1.00 million parameters, 60.94 million FLOPs, and a model size of 3.82 MB. DenseNet121, by contrast, requires 7.09 million parameters, 2896.12 million FLOPs, and 27.03 MB, yet achieves a lower weighted F1-score of 0.936±0.036. This is not a marginal efficiency advantage. It is a strong and unexpected reversal of the common assumption that larger networks naturally yield better results. In this task, the heaviest model is not the strongest. The result suggests that, when only about 270 training images are available per fold, representational excess does not translate into better generalization. What matters more is whether the architecture is well matched to the scale and structure of the data.

MobileNetV2 and EfficientNetB0 also perform strongly, but their efficiency profiles differ. MobileNetV2 offers the lowest latency among the top three baselines at 12.16 ms and the highest throughput at 82.24 images per second, making it especially attractive when rapid inference is prioritized. EfficientNetB0 reaches comparable predictive performance, but at a notably higher structural cost. This places MobileNetV3Small, MobileNetV2, and EfficientNetB0 in three distinct yet meaningful positions: MobileNetV3Small as the most compact, high-performing baseline; MobileNetV2 as the fastest, high-performing baseline; and

https://doi.org/10.31436/iiumej.v27i2.4341

EfficientNetB0 as the strongest, largest, lightweight baseline. Seen together, these results show that baseline selection in spinal X-ray classification should not be framed as a race for a single winner. It is more accurately a decision about which kind of efficiency-performance balance best fits the deployment objective.

MNASNet presents a different story. Although designed through neural architecture search with latency-aware objectives, it achieves only 0.905±0.029 in weighted F1-score, roughly 5 percentage points below the top baselines. Its inference speed is the fastest in the comparison, with a throughput of 87.17 images per second, but that advantage is not enough to offset the loss in diagnostic performance. In a clinical screening scenario, speed is valuable only if it does not come at too high a predictive cost. The result suggests that features optimized for generic mobile efficiency do not automatically transfer well to spinal radiographic morphology. In other words, being fast is not the same as being clinically sharp.

The comparison with prior studies further clarifies the contribution of the present benchmark. Earlier work on the same 338-subject dataset reported a mean three-class accuracy of 96.73% and a peak accuracy of 98.02% in [1], while [5] reported 99.01% accuracy. At first glance, these numbers may appear slightly higher than the cross-validation averages reported here. However, the comparison is not like-for-like. Those studies relied on fixed train-test splits, whereas the present work reports mean ± standard deviation under stratified 5-fold cross-validation. This difference is crucial. A single split can produce an impressive headline result, but it does not reveal how sensitive the model is to data partitioning. By contrast, the cross-validation protocol used here provides a more conservative and more credible estimate of generalization. Thus, the contribution of Table 4 is not merely another set of scores. It is a stronger answer to a more demanding question: not how well a model performs once, but how well it continues to perform as the evaluation becomes harder to game.

4.2. Proposed Models versus Baselines

Table 6 compares the proposed CBAM-enhanced and fusion-based models against the strongest baseline, MobileNetV3Small, under two training regimes: `class_weight` and `augmentation_only`. The central result is clear. Under `class_weight`, the proposed attention-enhanced models do not outperform the baseline. CBAM-MobileNetV3Small achieves a weighted F1-score of 0.951±0.024, which is slightly below the baseline MobileNetV3Small at 0.962±0.022. The same pattern appears in CBAM-MNASNet and, to a lesser extent, in CBAM-EfficientNetB0. Although these differences are modest and remain within approximately one standard deviation, their direction is consistent across architectures. This consistency is important because it shows that attention does not automatically translate into better generalization when the training data are limited.

This result is scientifically more useful than a simple improvement would have been. It shows that CBAM should not be treated as a universal performance booster. Under limited-data conditions, attention increases representational flexibility, but that added flexibility is only valuable if the training regime provides sufficient diversity to regularize it. When each fold contains only about 270 training samples, the model may gain more expressive capacity than the available supervision can reliably support. In that setting, the issue is not that attention fails in principle. Rather, the model becomes more adaptable than the data can safely guide.

The behavior of the fusion model reinforces the same point. Under `class_weight`, Fusion_MNv3_MNAS achieves a weighted F1-score of 0.951±0.033, again below the strongest baseline. Its performance converges toward that of its better branch, CBAM-MobileNetV3Small, rather than exceeding it. This behavior is consistent with the role of logit-

https://doi.org/10.31436/iiumej.v27i2.4341

level fusion. Fusion can reduce variance and soften the effect of a weaker branch, but it cannot create discriminative information that the branches themselves have not learned. Here, the weaker CBAM-MNASNet branch is partly compensated for, yet the overall gain remains limited because both branches are trained under the same restricted supervision regime. The result is therefore not surprising. Fusion improves reliability only when the underlying branch representations are sufficiently informative and sufficiently diverse.

Table 6. Proposed Models versus the Baseline under the `class_weight` and augmentation_only strategy

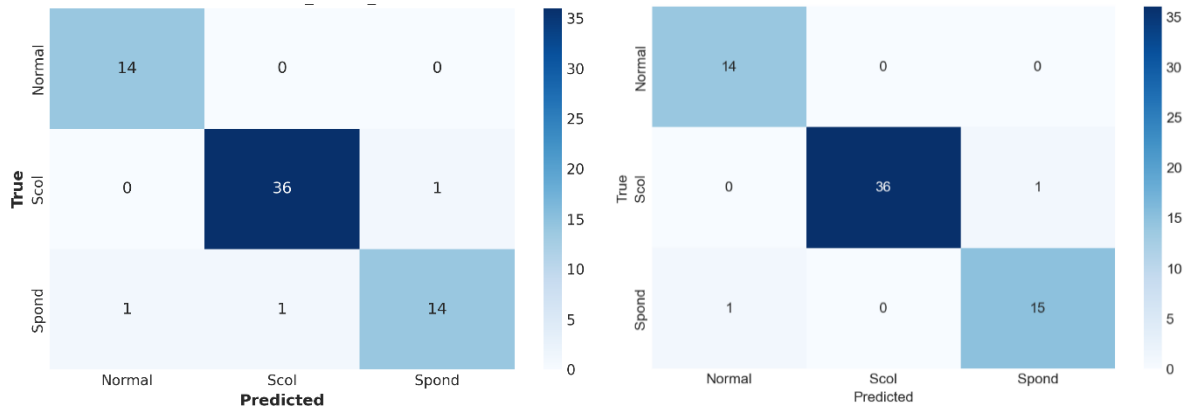
Model	`class_weight`		`augmentation_only`		Params (M)	Size (MB)	FLOPs (M)
	Accuracy	F1 (w)	Accuracy	F1 (w)			
MobileNetV3Small	0.962 ± 0.022	0.962 ± 0.022	0.953 ± 0.025	0.952 ± 0.026	1.001	3.819	61
CBAM_MobileNetV3Small	0.950 ± 0.024	0.951 ± 0.024	0.956 ± 0.026	0.955 ± 0.027	1.043	3.978	61
CBAM_MNASNet	0.891 ± 0.031	0.895 ± 0.028	0.917 ± 0.038	0.920 ± 0.036	3.472	13.243	336
CBAM_EfficientNetB0	0.959 ± 0.019	0.958 ± 0.019	0.947 ± 0.037	0.945 ± 0.038	4.377	16.696	414
Fusion_MNv3_MNAS	0.950 ± 0.034	0.951 ± 0.033	**0.976 ± 0.020**	**0.976 ± 0.020**	4.514	17.221	397

The performance changes under `augmentation_only`, and this is where the proposed framework becomes most revealing. Fusion_MNv3_MNAS reaches the highest weighted F1-score in the study, 0.976±0.020, clearly surpassing both the baseline MobileNetV3Small and all other proposed variants. This shift is the key empirical message of Table 5. The proposed models are not universally superior across all training conditions, but they become strongest when the training pipeline provides greater data diversity. Augmentation does more than rebalance the learning objective. It exposes the model to a broader range of anatomically plausible variations, thereby regularizing the additional parameters introduced by attention and fusion. Under this regime, the extra capacity of CBAM and dual-branch fusion is no longer a liability. It becomes useful.

This interaction between model design and training strategy is the main methodological insight of this section. Architecture alone does not determine performance in spinal X-ray classification. The same model can underperform under one regime and become the strongest performer under another. The implication is important: CBAM and fusion should not be evaluated in isolation, but in relation to the regularization environment in which they are trained. In the present study, augmentation-only training provides that environment more effectively than class weighting.

Figure 3 provides class-level evidence that supports this interpretation. The confusion matrices show that the remaining errors are concentrated mainly at the Normal-Scoliosis boundary. This is clinically plausible because mild scoliosis can resemble normal anatomical variation when curvature is subtle. By contrast, Spondylolisthesis is classified more consistently across models, indicating that its structural features are more visually distinct. The fusion model under `augmentation_only` reduces confusion at the Normal-Scoliosis boundary more effectively than the baseline under `class_weight`, suggesting improved sensitivity to subtle curvature-related patterns. Thus, the value of the proposed model is not merely that it raises an overall score, but that it improves discrimination in the most ambiguous part of the problem.

https://doi.org/10.31436/iiumej.v27i2.4341



(a) MobileNetV3Small under class_weight (b) Fusion_MNv3_MNAS under augmentation_only

Figure 3. Confusion Matrices (fold 5), in which residual errors concentrate on the Normal-Scoliosis boundary.

4.3. Efficiency Analysis

Figures 4 and 5 show that efficiency is not a side result in this study. It is part of the main result. When several models achieve similar predictive performance, the more important question becomes how much computation, memory, and time must be spent to obtain that performance. Viewed from this perspective, the trade-off space is highly structured rather than arbitrary.

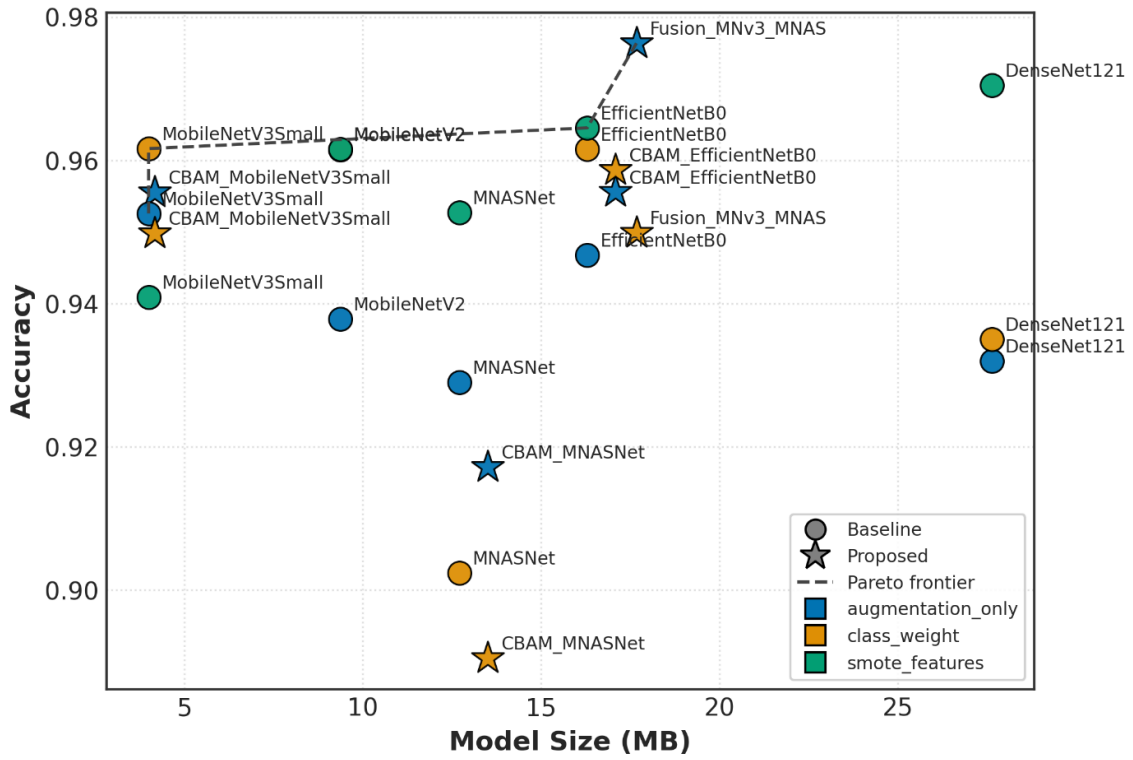


Figure 4. Model size (MB) versus accuracy for all evaluated models across three balancing strategies. The Pareto frontier highlights the smallest strong baseline, the intermediate trade-off region, and the highest-accuracy fusion configuration.

Figure 4 plots *accuracy versus model size* and reveals a compact Pareto frontier. At the smallest end of the spectrum, MobileNetV3Small remains the most storage-efficient strong

https://doi.org/10.31436/iiumej.v27i2.4341

baseline. With a model size of only 3.82 MB, it already achieves accuracy competitive with the best baseline results. At the other end, Fusion_MNv3_MNAS under `augmentation_only` achieves the highest accuracy in the study, but this gain comes at a substantially larger footprint of 17.22 MB. Between these two points, EfficientNetB0 occupies an intermediate position, offering a modest accuracy gain over the smallest model while increasing in size. The remaining models fall below this frontier and are therefore Pareto-inferior, meaning that they are larger without providing a compensating improvement in predictive performance. This is especially clear for DenseNet121, which is much larger than MobileNetV3Small yet does not deliver better accuracy on this dataset. The implication is direct: for small medical imaging datasets, more parameters do not automatically buy better generalization.

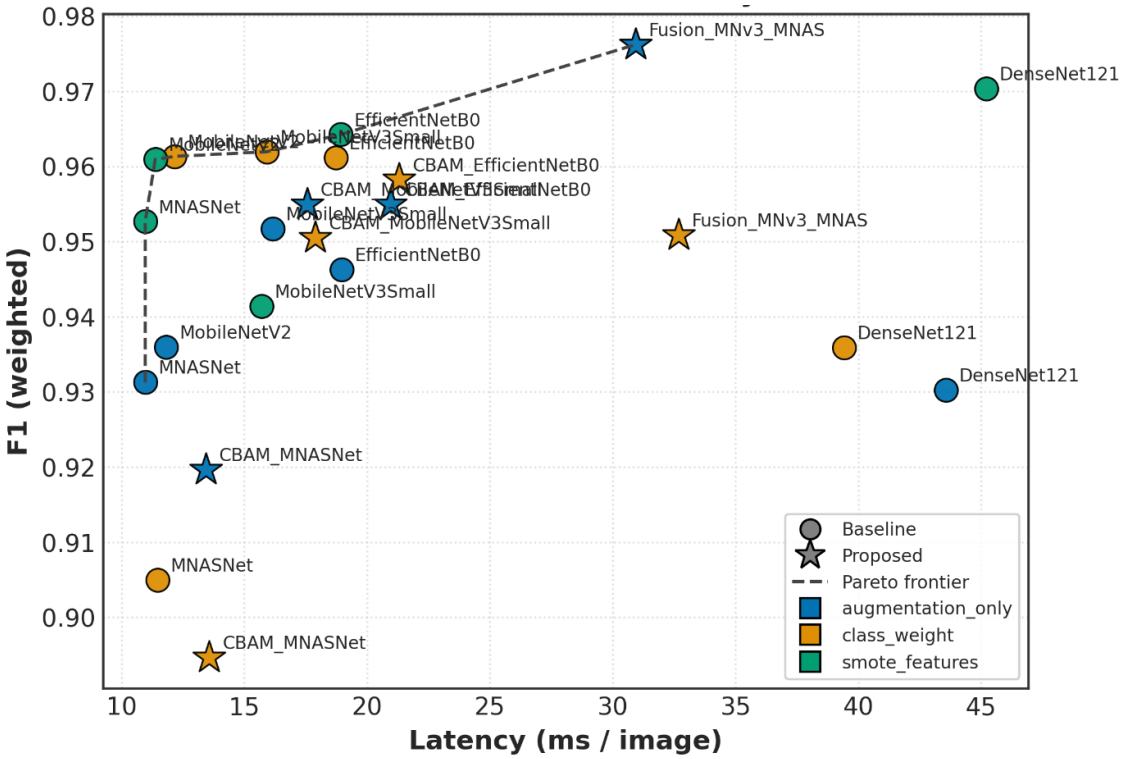


Figure 5. Inference latency (ms/image) versus weighted F1-score for all evaluated models across three balancing strategies. Lower latency is better. The figure highlights the trade-off between rapid inference and maximum predictive performance.

Figure 5 complements this view by plotting *weighted F1-score versus inference latency*, which is more relevant to deployment scenarios in which scans are processed individually. The figure shows that the best model depends on the operational constraint. If latency is the main concern, the most attractive region lies toward the left side of the plot, where lightweight baselines such as MobileNetV3Small and MobileNetV2 achieve strong weighted F1 Scores with relatively low inference times. MNASNet achieves the lowest overall latency, but its lower weighted F1-score makes that speed advantage less attractive in a clinical screening context. At the opposite end, Fusion_MNv3_MNAS under `augmentation_only` achieves the highest weighted F1-score, but with a clear latency penalty. This does not weaken the model's value. It simply defines its appropriate use case more precisely.

Taken together, Figures 4 and 5 show that no single model is optimal across all deployment scenarios. For memory-constrained or latency-sensitive settings, MobileNetV3Small remains

https://doi.org/10.31436/iiumej.v27i2.4341

the most practical choice because it combines strong predictive performance with very low storage cost. By contrast, when predictive performance is the dominant objective and a larger latency budget is acceptable, Fusion_MNv3_MNAS under `augmentation_only` becomes the preferred configuration because it extends the frontier upward. The broader lesson is that model selection in spinal X-ray classification should not be framed as a search for one universally best architecture. It should be framed as a matching problem between predictive ambition and operational constraint.

4.4. Effect of Class Balancing Strategy

Table 7 shows that model performance in this study is shaped not only by architecture, but also by how class imbalance is handled during training. This is a central result of the paper. Among the three strategies, `augmentation_only` produces the highest overall performance, with Fusion_MNv3_MNAS achieving a weighted F1-score of 0.976±0.020, the best result in the entire experimental matrix. More importantly, this is the only regime in which the proposed CBAM-enhanced and fusion-based models consistently outperform their baseline counterparts. The implication is clear: attention and fusion do not become effective simply because they are added to the architecture. They become effective when the training pipeline supplies enough data diversity to regularize the added representational capacity.

Table 7. Best-performing model under each class balancing strategy (5-fold stratified cross-validation, mean ± standard deviation)

Strategy	Best Model	Accuracy	F1 (weighted)	Key Characteristic
class_weight	MobileNetV3Small	0.962 ± 0.022	0.962 ± 0.022	Weighted cross-entropy, no synthetic data
smote_features	DenseNet121	0.971 ± 0.010	0.970 ± 0.011	SMOTE on 1024-D embeddings
augmentation_only	Fusion_MNv3_MNAS	0.976 ± 0.020	0.976 ± 0.020	Strong augmentation (geometric + photometric)

This helps explain why `augmentation_only` performs best. Class weighting changes the optimization objective, but it does not increase the visual diversity of the training data. Augmentation does. By exposing the model to a broader range of anatomically plausible geometric and photometric variation, it acts as a form of data-driven regularization. Under this regime, the CBAM module is better able to learn meaningful channel-spatial relationships instead of fitting too closely to a limited set of images. In that sense, augmentation does more than mitigate imbalance. It improves the match between the model's flexibility and the diversity of the supervision signal.

By contrast, `class_weight` provides a strong and stable baseline rather than the highest score. MobileNetV3Small achieves a weighted F1-score of 0.962±0.022 under this strategy, indicating that loss reweighting alone is sufficient to achieve competitive performance. The difference between the best `class_weight` result and the best `augmentation_only` result is approximately 1.4 percentage points in weighted F1-score. Given the cross-validation variability observed in this study, this improvement is consistent but not strongly separated at the present dataset scale. From a practical perspective, this makes `class_weight` with MobileNetV3Small the most conservative and deployment-ready configuration: it is simple, reproducible, lightweight, and strong enough to remain highly competitive without requiring a more aggressive training pipeline.

https://doi.org/10.31436/iiumej.v27i2.4341

The `smote_features` strategy produces a different pattern and must be interpreted more carefully. Its best result is obtained by DenseNet121, which improves from 0.936 to 0.970 in weighted F1-score. This suggests that feature-space oversampling can be effective for compensating class imbalance when the classifier is trained on fixed backbone embeddings. However, this setting is not directly comparable to the end-to-end strategies. In the SMOTE pipeline, oversampling is applied only after feature extraction, and only the classifier head is trained. As a result, the learning problem is structurally simpler and computationally much cheaper than full end-to-end optimization. For that reason, `smote_features` should be interpreted as a feature-level baseline rather than as an equivalent alternative to the end-to-end balancing strategies.

This difference also explains why SMOTE is not evaluated on the proposed CBAM and fusion models. Feature-space SMOTE operates on flattened embeddings after global average pooling, whereas CBAM modifies intermediate feature maps in the (C, H, W) space before pooling. There is therefore no single post-attention embedding that can be used for SMOTE interpolation without changing the learning pipeline itself. Applying SMOTE after CBAM would require retraining the model end-to-end, which defeats the purpose of the two-stage feature-space approach. Applying it directly in feature-map space would violate the assumptions of the SMOTE algorithm. Accordingly, the `smote_features` strategy is structurally incompatible with the CBAM-based and fusion-based models. This limitation is methodological rather than incidental, and the comparison must be interpreted with that constraint in mind.

Taken together, Table 7 supports a broader conclusion: class balancing strategy is not a secondary implementation detail. It changes which model becomes strongest. Under `class_weight`, the most convincing result comes from a simple lightweight baseline. Under `augmentation_only`, the proposed fusion architecture achieves the best performance in the study. The main lesson is therefore not that one balancing strategy is universally superior, but that the effect of balancing depends on the architecture it supports and the kind of learning pipeline it enables.

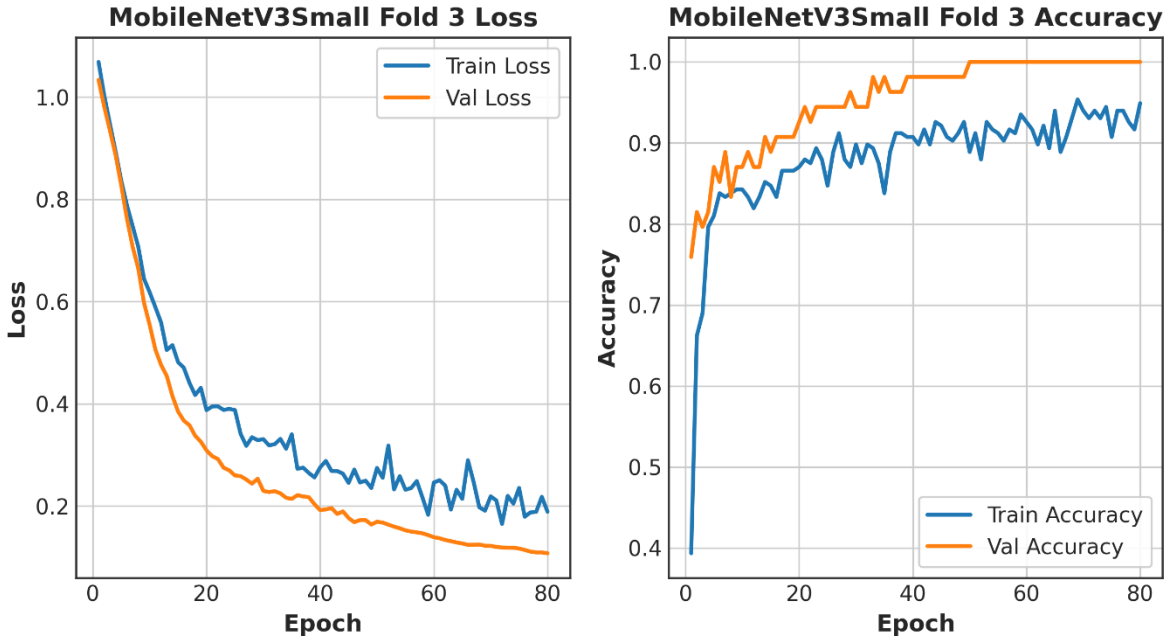
4.5. Learning Curves and Convergence

Figure 6 shows that the convergence behavior in this study is not random or unstable. It follows a clear pattern shaped by the two-phase training protocol. In *Phase 1*, where the backbone remains frozen and only the classifier head is trained at a learning rate of 1×10^{-4}, both models converge rapidly. Validation performance stabilizes within roughly 15 to 20 epochs, indicating that the pretrained ImageNet features already provide a strong starting representation for spinal X-ray classification. The early plateau, followed by one or two learning-rate reductions, suggests that optimization reaches a stable region without requiring prolonged training. This supports the design of the first phase: freezing the backbone allows efficient adaptation while limiting unnecessary parameter updates on a small dataset.

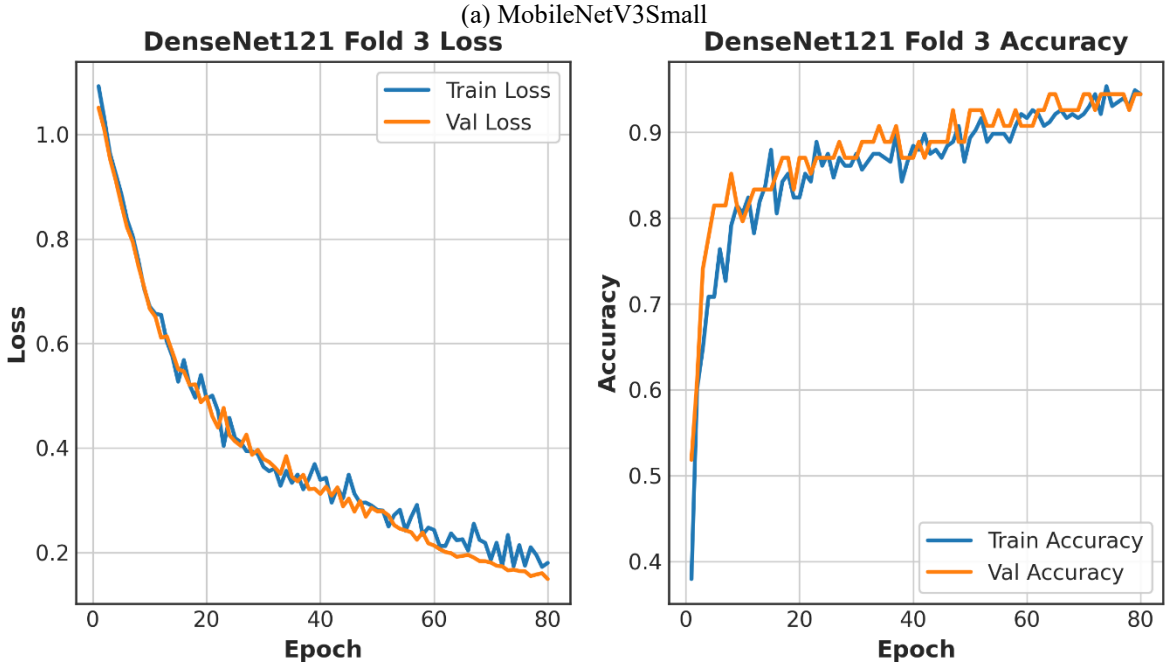
The transition to *Phase 2*, in which a subset of higher-level layers is unfrozen, and the learning rate is reduced to 1×10^{-5}, reveals a more informative contrast. For lightweight models such as MobileNetV3Small, the gain from fine-tuning is small, and the validation curve remains largely flat after unfreezing. This suggests that the pretrained features are already well aligned with the target task and that additional adaptation yields only limited benefit. In contrast, higher-capacity models, including EfficientNet-based and fusion models, show modest but more visible improvements after fine-tuning, typically on the order of 0.5 to 1 percentage point in weighted F1-score. The implication is that fine-tuning is not universally

https://doi.org/10.31436/iiumej.v27i2.4341

beneficial. Its value depends on how much unused representational capacity remains in the model and how much task-specific refinement the pretrained features still require.



(a) MobileNetV3Small



(b) DenseNet121

Figure 6. Learning curves under class_weight (fold 3). The dashed vertical line marks the Phase 1 to Phase 2 boundary. DenseNet121 exhibits clear train-validation divergence after epoch 30.

DenseNet121 provides the clearest counterexample to the assumption that more capacity is always useful. Its learning curve shows a noticeable divergence between training and validation losses beginning around epoch 30: training loss continues to decrease, while validation loss begins to rise. This is the characteristic signature of overfitting. Given the model's relatively large parameter count compared with the limited number of training samples per fold, this behavior is not surprising. More importantly, it provides direct empirical support for the broader conclusion of this paper: on small medical imaging datasets, larger architectures

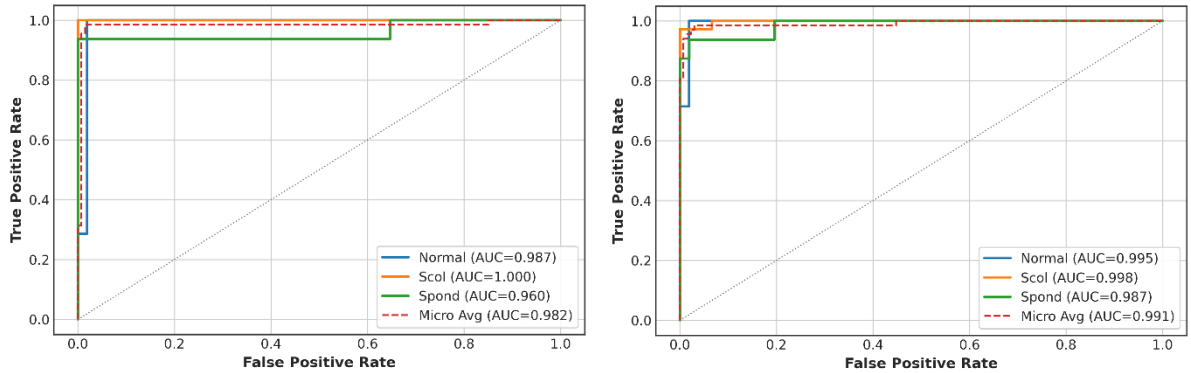
https://doi.org/10.31436/iiumej.v27i2.4341

can fit the training set more easily but generalize worse. In this setting, early stopping is not merely a routine safeguard. It functions as an active regularizer, preventing further training from degrading validation performance.

Equally important is what the learning curves do not show. None of the models exhibits unrealistically fast convergence to near-perfect training performance, which would raise concern about data leakage, trivial shortcuts, or overly aggressive preprocessing. Instead, the curves remain smooth and progressively saturating. This pattern supports the integrity of the experimental setup and suggests that the reported performance is driven by learned feature discrimination rather than accidental artifacts. Taken together, Figure 6 shows that the proposed training protocol achieves stable convergence, controlled generalization, and effective regularization, while also reinforcing a central result of the study: lightweight architectures offer the most favorable balance between learnability and generalization on this dataset.

4.6. ROC Analysis

Figure 7 provides a threshold-independent view of model behavior and confirms the main interpretation drawn from the accuracy and weighted F1 results. All high-performing models, particularly those with weighted F1-scores above 0.95, also achieve macro-average AUC values of approximately 0.98 or higher. This agreement is important because it shows that their performance is not tied to a single operating threshold. Instead, the models are learning class representations that remain well separated across a wide range of decision boundaries. In practical terms, the earlier observed performance gains reflect genuine improvements in discrimination rather than threshold-specific optimization.



(a) MobileNetV3Small (class_weight) (b) Fusion_MNv3_MNAS (augmentation_only)

Figure 7. Per-class one-vs-rest ROC curves. Spondylolisthesis achieves the highest per-class AUC across both models.

The ROC profiles of MobileNetV3Small under `class_weight` and Fusion_MNv3_MNAS under `augmentation_only` are especially informative. Their curves are both strong, but the fusion model shows a slightly more favorable separation, consistent with its higher weighted F1-score. This supports the interpretation that the gain from the fusion model is real but selective: it does not arise from arbitrary threshold tuning, but from improved class separability under a more suitable regularization regime.

At the class level, Spondylolisthesis remains the easiest category to distinguish, with per-class AUC values approaching 0.99 across the strongest models. This result is consistent with the confusion-matrix analysis and is clinically plausible. Spondylolisthesis involves a relatively

https://doi.org/10.31436/iiumej.v27i2.4341

pronounced geometric displacement of vertebral bodies, which produces a stronger global structural cue than the more subtle curvature changes seen in early or mild scoliosis. Such macroscopic patterns are more likely to survive resizing and to remain accessible to convolutional features transferred from large-scale pretraining.

The main residual difficulty remains at the Normal-Scoliosis boundary. The ROC curves for these two classes remain slightly weaker than those of Spondylolisthesis, indicating that the task is limited less by gross pathology detection than by discrimination of subtle borderline cases. This again aligns with the earlier confusion-matrix analysis. Mild scoliosis may preserve large regions of apparently normal anatomy, making the distinction dependent on finer geometric cues that are harder to preserve at standard input resolution. Thus, the ROC analysis does not introduce a new conclusion. It strengthens an existing one: the core challenge in this dataset is not recognizing obvious abnormality, but separating normal spines from mild curvature patterns with high reliability.

4.7. Summary of Key Findings

The results of this study converge on four main findings. *First*, among all evaluated backbones, MobileNetV3Small is the strongest efficiency-oriented baseline. It achieves top-tier predictive performance with only 1.0 million parameters and 61 million FLOPs, showing that, on this dataset, lightweight design is not a compromise but an advantage. The comparison with DenseNet121 makes this especially clear: substantially greater model capacity does not translate into better generalization when training samples are limited.

Second, the effectiveness of the proposed CBAM-enhanced and fusion-based models is conditional rather than universal. Under `class_weight`, they do not outperform their plain-backbone counterparts. Under `augmentation_only`, however, Fusion_MNv3_MNAS becomes the best-performing configuration in the study, achieving a weighted F1-score of 0.976. The implication is that attention and fusion are not inherently beneficial in isolation. Their value emerges when the training regime provides enough data diversity to regularize the added representational capacity.

Third, `augmentation_only` is the most effective end-to-end balancing strategy in this study. Unlike `smote_features`, it operates within the same training pipeline as the proposed architectures and does not rely on a separate feature-level oversampling stage. More importantly, it consistently improves models with the greatest representational flexibility, indicating that data diversity is a more effective enabler of attention and fusion than loss reweighting alone. In this setting, augmentation functions not only as a balancing tool, but also as the key regularizer that allows more expressive architectures to generalize.

Finally, all analyses converge on the same residual source of error: the Normal–Scoliosis boundary. Spondylolisthesis is consistently well separated, while mild scoliosis remains difficult to distinguish from normal anatomy. The remaining challenge in spinal X-ray classification is therefore not the detection of obvious pathology but the reliable recognition of subtle curvature-related variation.

These findings are based on a single public dataset of 338 subjects. Although stratified 5-fold cross-validation strengthens the estimate relative to a fixed split, generalization to broader clinical populations, different acquisition settings, or more diverse radiographic characteristics remains to be established. Accordingly, the contribution is best understood as a robust, efficiency-aware methodological benchmark rather than a final deployment claim.

https://doi.org/10.31436/iiumej.v27i2.4341

5. ABLATION STUDIES

The main results presented in Section 4 identify which models perform best. The ablation studies explain why. This distinction is important because the proposed framework combines several interacting design choices, including attention, fine-tuning, fusion, and class balancing. Without controlled ablation, an observed gain could be incorrectly attributed to the wrong component. For that reason, four targeted ablation groups are conducted, each varying one factor while holding the others fixed. The purpose is not to generate more tables, but to isolate which design choices genuinely contribute to performance, under which training conditions, and with what practical implications. In this study, the ablations are especially important because the main results already suggest that performance is governed by model–training interaction rather than architecture alone.

5.1. CBAM Sub-Module Analysis

The first ablation examines a simple but necessary question: *if attention helps, which part of attention is most important*? To answer this, the contribution of the individual CBAM components is isolated on the MobileNetV3Small backbone under the `class_weight` strategy. Four controlled conditions are evaluated: the plain backbone without attention, channel attention only, spatial attention only, and full CBAM with sequential channel-spatial attention. This design allows the effect of each attention component to be interpreted directly rather than inferred indirectly from the full model.

Table 8. CBAM sub-module ablation on MobileNetV3Small under the class_weight strategy (5-fold stratified cross-validation, mean ± standard deviation).

Condition	Accuracy	F1 (weighted)	Params
No attention (baseline)	0.965 ± 0.022	0.965 ± 0.022	1,001,251
Channel attention only	0.953 ± 0.019	0.954 ± 0.019	1,042,723
Spatial attention only	0.956 ± 0.027	0.957 ± 0.027	1,001,349
Full CBAM	0.953 ± 0.019	0.954 ± 0.019	1,042,821

The result shown in Table 8 is unambiguous. All three attention variants underperform the plain baseline. The baseline achieves the highest weighted F1-score at 0.965±0.022, followed by spatial attention only at 0.957±0.027, while both channel attention only and full CBAM reach 0.954±0.019. The full CBAM configuration, therefore, reduces weighted F1 by about 1.1 percentage points relative to the baseline while also adding approximately 41,000 parameters. This is a small architectural increase, but it produces no compensating gain. More importantly, the pattern is systematic rather than random: the ranking is consistent enough to suggest that, under `class_weight` training on this dataset, the added attention capacity is not translated into better generalization.

This is an important negative result. It challenges the common expectation that attention modules are inherently beneficial once inserted into a CNN backbone. In the present setting, that assumption does not hold. The issue is not that CBAM is poorly designed. The issue is that the training regime does not provide enough data diversity to support the added representational flexibility. Under these conditions, the model gains more freedom than the supervision signal can reliably guide.

Among the three attention variants, spatial attention performs better than channel attention, although the margin is modest. This difference is meaningful. Spinal disorder classification depends primarily on geometric structure, especially curvature in scoliosis and vertebral

https://doi.org/10.31436/iiumej.v27i2.4341

displacement in spondylolisthesis. Spatial attention is more naturally aligned with this kind of signal because it emphasizes where discriminative evidence occurs in the image. Channel attention, by contrast, is more indirect in this context because it reweights feature channels without explicitly sharpening spatial localization. The result, therefore, suggests that, when attention is useful for this task, spatial selectivity is more relevant than channel recalibration.

At the same time, spatial attention alone still does not surpass the plain baseline. This shows that partial alignment with the task is not sufficient on its own. The broader conclusion is that, under the `class_weight` regime, the full CBAM module is not justified for MobileNetV3Small on this dataset. Channel and spatial attention do not combine to produce an additive benefit, and the plain backbone remains the strongest option.

This should not be interpreted as a rejection of attention mechanisms in general. Rather, it provides direct evidence that the value of attention is conditional. As shown later in Section 5.4, the same attention-based models behave differently under `augmentation_only`, where richer data diversity allows the added representational capacity to be regularized more effectively. The practical lesson is therefore precise: in spinal X-ray classification, attention should not be evaluated as a universal architectural upgrade, but as a component whose usefulness depends on the training regime that supports it.

5.2. Fine-Tuning Depth

This ablation examines a practical question in transfer learning: *how much of the pretrained backbone should be unfrozen before additional adaptation stops helping*? The question matters because too little fine-tuning leaves the model dependent on generic pretrained features, whereas too much fine-tuning risks unnecessary complexity and instability on a small medical dataset. To study this trade-off directly, five fine-tuning fractions were evaluated on CBAM_MobileNetV3Small under the `class_weight` strategy, ranging from a fully frozen backbone to unfreezing half of the network. This provides a controlled view of how increasing trainable capacity translates into measurable gains.

Table 9. Effect of backbone fine-tuning depth on CBAM_MobileNetV3Small under the class_weight strategy (5-fold stratified cross-validation, mean ± standard deviation).

Fraction Unfrozen	Trainable Params	Accuracy	F1 (weighted)	Train Time (s)
0% (frozen)	115,813	0.941 ± 0.018	0.942 ± 0.017	71.2
10%	409,909	0.950 ± 0.025	0.951 ± 0.024	107.3
20% (default)	704,581	0.956 ± 0.021	0.956 ± 0.020	111.1
30%	838,477	0.962 ± 0.022	0.962 ± 0.022	113.2
50%	904,109	0.962 ± 0.022	0.962 ± 0.022	116.0

The pattern is clear, as shown in Table 9. Performance improves steadily as the fine-tuning fraction increases from 0% to 30%, then levels off. Both the 30% and 50% settings produce identical mean accuracy, identical weighted F1-score, and identical variance across folds. This indicates that the useful domain-specific adaptation is already captured by unfreezing about 30% of the backbone. Beyond that point, additional trainable parameters do not yield further gain. They slightly increase optimization cost, but not learning quality.

Figure 8 makes this pattern easy to interpret. The fully frozen setting performs worst, with a weighted F1-score of 0.942±0.017, which is about 2 percentage points below the 30% plateau. This gap is large enough to show that feature extraction alone is insufficient for this task. Even with strong ImageNet pretraining, some degree of domain-specific adaptation is necessary to

https://doi.org/10.31436/iiumej.v27i2.4341

capture the geometric characteristics of spinal disorders. In this dataset, fine-tuning is therefore not optional. It is part of what makes the model competitive.

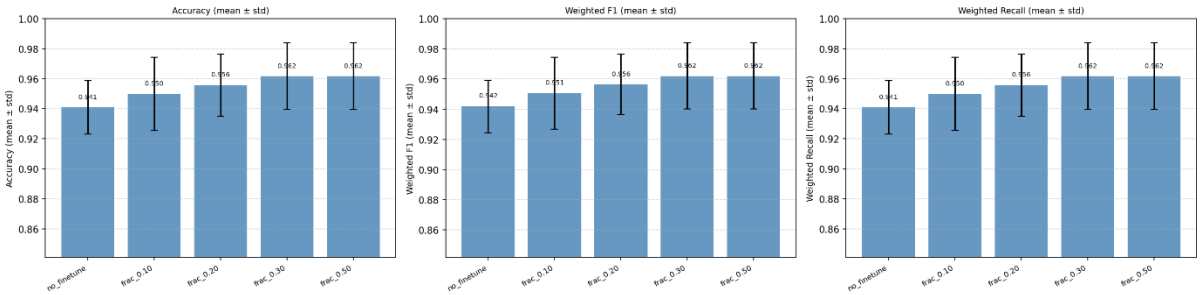


Figure 8. Weighted F1-score as a function of the backbone fine-tuning fraction. Performance plateaus at 30%, with the 30% and 50% conditions producing identical per-fold metrics.

At the same time, the results do not support the opposite extreme either. Unfreezing more than 30% of the backbone provides no measurable advantage. The 50% setting increases training time relative to 30%, but produces the same performance. The implication is practical: partial fine-tuning should continue only until performance saturates. Beyond that point, additional flexibility is unnecessary.

This result also clarifies the reviewer's concern regarding the difference between the benchmark protocol and the ablation optimum. In the main comparative experiments, 20% unfreezing was used as the *default benchmark setting* to maintain a uniform training protocol across all evaluated models. The purpose of that setting was fairness of comparison, not model-specific optimization. The ablation study shows that, for CBAM_MobileNetV3Small, a 30% fraction is the better task-specific choice, improving weighted F1 from 0.956 to 0.962 with only a negligible increase in training time. The main results should therefore be interpreted as results under a controlled common protocol, whereas this ablation identifies the more refined optimum for this specific model.

An equally important observation is what does not happen. Contrary to the common concern that unfreezing too many layers on a small dataset will inevitably lead to overfitting, no performance collapse is observed even at 50% unfreezing. This suggests that the training protocol, especially early stopping and adaptive learning rate scheduling, is effective in constraining the optimization process. In this setting, generalization is shaped not only by the parameter count but also by how the fine-tuning process is controlled.

Taken together, the fine-tuning ablation yields a simple design rule: a fully frozen backbone is too rigid, while deeper fine-tuning beyond 30% is unnecessary. For CBAM_MobileNetV3Small on this dataset, the useful adaptation window lies between. That is the point at which pretrained knowledge is preserved, but the model is given enough freedom to learn the subtle geometric signatures of spinal pathology.

5.3. Fusion versus Individual Branches

This ablation tests a central design question of the proposed framework: *does logit-level fusion between two CBAM-enhanced backbones yield a model stronger than either branch alone*? The motivation is straightforward. MobileNetV3Small and MNASNet differ in width, embedding dimension, and design objective, so they may learn complementary representations of spinal anatomy. If that complementarity is real, combining their predictions should improve robustness and, potentially, average performance. To isolate the effect of fusion, the two

individual branches are compared directly against the fused model using the `class_weight` strategy.

Table 10. Fusion versus individual CBAM branches under the class_weight strategy (5-fold stratified cross-validation, mean ± standard deviation).

Model	Accuracy	F1 (weighted)	Params (M)	Size (MB)
CBAM_MobileNetV3Small	0.944 ± 0.019	0.945 ± 0.019	1.040	3.980
CBAM_MNASNet	0.920 ± 0.042	0.922 ± 0.040	3.470	13.240
Fusion_MNv3_MNAS	0.941 ± 0.029	0.943 ± 0.028	4.510	17.220

The result is instructive, as shown in Table 10. Under `class_weight`, the fusion model does not outperform its strongest branch. Fusion_MNv3_MNAS achieves a weighted F1-score of 0.943±0.028, slightly below CBAM_MobileNetV3Small at 0.945±0.019. The gap is small and remains within one standard deviation, but the direction is clear. Fusion does not create a stronger classifier when one branch is consistently weaker than the other. Instead, uniform logit averaging pulls the final prediction toward the mean behavior of the two branches. In this case, the weaker CBAM_MNASNet branch reduces the advantage of the stronger CBAM_MobileNetV3Small branch rather than contributing enough complementary information to offset its own weakness.

At the same time, fusion provides something different from a higher mean score: it improves stability. CBAM_MNASNet shows substantial fold-to-fold variability, with a weighted F1 standard deviation of about 4.0% and worst-fold performance at 0.858. The fusion model compresses that variability and raises the worst-fold outcome to 0.901. This means that, even without improving the average score, fusion reduces the risk of particularly poor runs. That property is practically relevant. In clinical decision-support settings, a model that is slightly less optimal on average but more resistant to large performance drops may still be preferable, especially when reliability matters more than chasing a marginal gain in peak accuracy.

This result clarifies the role of fusion in the present framework. Under limited regularization, fusion primarily serves as a stabilizer rather than a performance amplifier. It can soften the effect of a weaker branch, but it cannot recover information that the branches themselves have not learned. For fusion to produce a genuine accuracy gain, both branches must contribute useful and sufficiently distinct representations.

That is exactly what appears under `augmentation_only`. In that regime, Fusion_MNv3_MNAS reaches a weighted F1-score of 0.976, clearly surpassing both individual branches. The contrast is important. It shows that fusion benefits from branch diversity only when both branches are trained under conditions that allow their representations to generalize well. Without sufficient regularization, the weaker branch contributes more noise than complementarity. With stronger augmentation, that same branch becomes more useful, and fusion begins to realize its intended benefit.

From a deployment perspective, the fusion model remains within practical limits despite its added complexity. With 4.51 million parameters and a model size of 17.22 MB, it is still substantially lighter than DenseNet121. Thus, the cost of fusion is real but controlled. The main conclusion of this ablation is therefore precise: *fusion is not guaranteed to improve mean performance under all training conditions, but it does improve stability and can become the best-performing strategy when the branches are sufficiently regularized*. In this study, the value of fusion is conditional rather than universal, consistent with the broader pattern observed in attention and class balancing.

https://doi.org/10.31436/iiumej.v27i2.4341

5.4. Balancing Strategy Interaction with Proposed Models

This ablation tests a question that is easy to overlook but central to the interpretation of the entire study: *Does the best model remain the best when the balancing strategy changes*? If the ranking changes, then architecture alone is not the real explanation of performance. To examine this directly, four proposed models were evaluated under both `class_weight` and `augmentation_only`. This allows the effect of the training regime to be separated from the effect of the architecture.

As shown in Table 11, the ranking is clearly not stable across strategies. Under class_weight, CBAM_EfficientNetB0 achieves the highest weighted F1-score at 0.958±0.019, while Fusion_MNv3_MNAS reaches only 0.948±0.022. Under augmentation_only, the order changes completely: Fusion_MNv3_MNAS becomes the strongest model at 0.974±0.026, whereas CBAM_EfficientNetB0 drops to 0.946±0.038. This inversion is the most important result of the ablation. It shows that conclusions about model superiority are conditional on the training regime. A model that appears suboptimal under one balancing strategy may become the best model under another.

Table 11. Performance of proposed models under class_weight and augmentation_only strategies (5-fold stratified cross-validation, mean ± standard deviation).

Model	Strategy	Accuracy	F1 (weighted)
CBAM_EfficientNetB0	class_weight	0.959 ± 0.019	0.958 ± 0.019
CBAM_EfficientNetB0	augmentation_only	0.947 ± 0.037	0.946 ± 0.038
CBAM_MobileNetV3Small	class_weight	0.956 ± 0.023	0.956 ± 0.023
CBAM_MobileNetV3Small	augmentation_only	0.947 ± 0.023	0.946 ± 0.023
CBAM_MNASNet	class_weight	0.902 ± 0.038	0.906 ± 0.036
CBAM_MNASNet	augmentation_only	0.926 ± 0.043	0.928 ± 0.041
Fusion_MNv3_MNAS	class_weight	0.947 ± 0.022	0.948 ± 0.022
Fusion_MNv3_MNAS	augmentation_only	0.973 ± 0.026	0.974 ± 0.026

The models separate into two distinct response patterns. CBAM_MNASNet and Fusion_MNv3_MNAS improve substantially under `augmentation_only`, gaining about 2.2 to 2.6 percentage points in weighted F1-score. By contrast, CBAM_EfficientNetB0 and CBAM_MobileNetV3Small decline slightly. This divergence is not random. The two stronger models under `class_weight` are already operating near the performance ceiling of this dataset, leaving limited room for further gains from augmentation. Additional input variability can then act more as a perturbation than as a benefit. In contrast, the weaker and higher-capacity models have more headroom and benefit more from the extra diversity introduced by augmentation.

This explains why `augmentation_only` works best for the proposed fusion architecture. Fusion_MNv3_MNAS has the largest representational capacity among the proposed models and therefore also the greatest need for regularization. Under `class_weight`, that capacity is not fully used. Under `augmentation_only`, the model is exposed to a broader distribution of anatomically plausible inputs, thereby reducing variance and enabling the two branches to learn more useful, complementary representations. The gain is therefore not caused by augmentation in isolation, but by the match between augmentation and model capacity. In this study, augmentation is the condition that allows fusion to become effective.

The same interaction also clarifies why `smote_features` is excluded from this comparison. SMOTE operates on flattened feature embeddings after global average pooling, whereas CBAM modifies intermediate feature maps in the (C, H, W) space before pooling. As

https://doi.org/10.31436/iiumej.v27i2.4341

a result, there is no single feature representation that both preserves the effect of CBAM and remains suitable for SMOTE interpolation. Applying SMOTE after CBAM would require retraining the full model end-to-end, defeating the purpose of the two-stage feature-space pipeline. Applying it directly to feature maps is not supported by the SMOTE formulation. The exclusion of SMOTE for CBAM and fusion models is therefore a structural constraint of the method, not an omitted experiment.

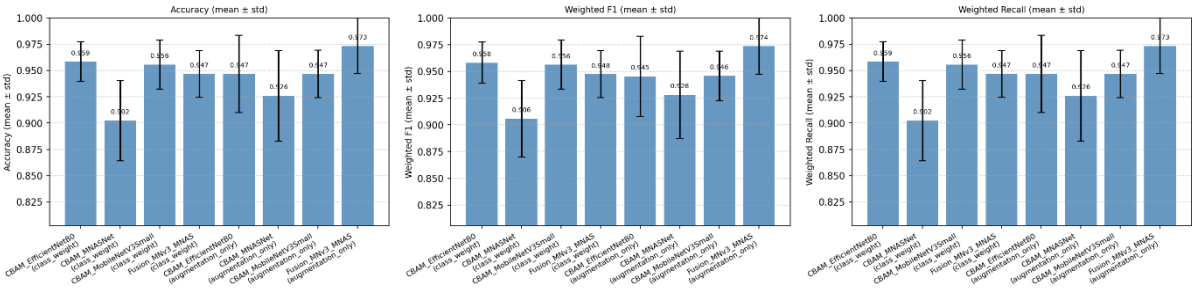


Figure 9. Weighted F1-score for each proposed model under class_weight and augmentation_only strategies. Fusion_MNv3_MNAS exhibits the largest strategy-dependent gain (+2.6 pp F1)

Figure 9 makes the strategy dependence visually explicit. The largest gain belongs to Fusion_MNv3_MNAS, while the two strongest single-branch models under `class_weight` show slight decreases under `augmentation_only`. This confirms that the contribution of CBAM and fusion is real but conditional. Evaluating these models only under `class_weight` would lead to the misleading conclusion that attention and fusion are ineffective. The more accurate conclusion is narrower and more useful: *on this dataset, CBAM and fusion become advantageous only when paired with sufficient data-driven regularization*. For the proposed models, `augmentation_only` is therefore the most appropriate training regime.

5.5. Summary of Ablation Findings

The ablation studies lead to two clear conclusions. *First*, attention is not universally beneficial in small-data settings. Under the class_weight regime, all CBAM variants underperform the plain backbone, with spatial attention proving the strongest variant. This is consistent with the geometric nature of the task: curvature and vertebral displacement favor spatial localization over channel reweighting. The fine-tuning ablation adds a complementary constraint: a fully frozen backbone is too rigid, while unfreezing beyond 30% yields no further gain. Useful domain adaptation is necessary, but bounded.

Second, and more importantly, model performance is governed by the interaction between architecture and training strategy rather than by architecture alone. Under `class_weight`, the proposed fusion model does not outperform its strongest branch. Under `augmentation_only`, the same model becomes the best-performing configuration in the study, with a weighted F1-score gain of about 2.6 percentage points. This directly explains why `augmentation_only` performs best: it provides the data diversity required to regularize the added capacity introduced by attention and fusion. The methodological implication is that, on small medical imaging datasets, architectural design and training strategy must be optimized jointly. Evaluating one without the other can mislead conclusions about which model is actually strongest.

6. CONCLUSIONS

This study presented an efficiency-aware deep learning framework for three-class classification of spinal disorders from X-ray images. The central message is that, on a dataset of this scale, the strongest model is not the largest, but the one whose architecture, regularization, and training strategy are matched to the task. MobileNetV3Small reached a weighted F1-score of 0.962 with only 1.0 million parameters and 61 million FLOPs, surpassing far heavier baselines such as DenseNet121 while remaining substantially smaller and faster.

CBAM and dual-branch fusion were not universally beneficial. Under `class_weight`, they did not surpass plain backbones, but under `augmentation_only`, `Fusion_MNv3_MNAS` achieved the best overall weighted F1-score of 0.976. Ablation analyses clarified the underlying mechanism: attention requires sufficient regularization to be useful, fusion improves stability before mean accuracy, and partial fine-tuning is necessary but saturates near 30% of the backbone. Throughout ROC, confusion-matrix, and efficiency analyses, Spondylolisthesis remained reliably separable, while the dominant residual difficulty lay at the Normal–Scoliosis boundary.

The principal contribution is therefore methodological rather than a single headline number: in spinal X-ray classification, lightweight models can be remarkably strong, architectural innovations are conditional rather than absolute, and reliable progress comes from co-optimizing model design with the training regime that allows it to generalize. Future work will extend this benchmark to larger, multi-center datasets and to additional spinal pathologies, to test how far the present conclusions generalize beyond the 338-subject setting examined here.

ACKNOWLEDGMENT

This research was supported by the SASMEC @ IIUM Research Grant (Project ID: SRG25-098-0098). The second author gratefully acknowledges the Kulliyyah of Engineering IIUM Engineering Merit Scholarship 2025 (KOEIEMS25). Ethical approval for this study, including the collection of additional scoliosis X-ray and ultrasound images, was granted by the IIUM Research Ethics Committee (IREC) under application number IREC-2025-3. The authors also wish to note that Figure 2 has been submitted for copyright registration to the Malaysian Intellectual Property Office through the IIUM Library. The authors used generative AI tools (ChatGPT and Claude) solely for language refinement and structural editing during manuscript preparation; all AI-assisted text was reviewed and verified by the authors, and no AI was used to generate research data, results, or substantive technical content.

REFERENCES

[1] M. Fraiwan, Z. Audat, L. Fraiwan, and T. Manasreh, "Using deep transfer learning to detect scoliosis and spondylolisthesis from x-ray images," *Plos One,* vol. 17, no. 5, p. e0267851, 2022.

[2] R. R. Maaliw III, "SCOLIONET: an automated scoliosis Cobb angle quantification using enhanced X-ray images and deep learning models," *Journal of Imaging,* vol. 9, no. 12, p. 265, 2023.

[3] M. I. Jamaludin, T. S. Gunawan, R. K. Karupiah, S. A. Zabidi, M. Kartiwi, and Z. Zakaria, "Optimizing U-Net Architecture with Feed-Forward Neural Networks for Precise Cobb Angle Prediction in Scoliosis Diagnosis," *Indonesian Journal of Electrical Engineering and Informatics (IJEEI),* vol. 11, no. 3, pp. 883-895, 2023.

[4] N. A. Makhdoomi *et al.*, "Development of scoliotic spine severity detection using deep learning algorithms," in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, 2022: IEEE, pp. 0574-0579.

[5] H. Güneş, C. Hark, and A. E. Akkaya, "Comparison of deep learning models and optimization algorithms in the detection of scoliosis and spondylolisthesis from X-ray images," *Sakarya University Journal of Science,* vol. 28, no. 2, pp. 438-451, 2024.

[6] G. M. Trinh *et al.*, "Detection of lumbar spondylolisthesis from X-ray images using deep learning network," *Journal of Clinical Medicine,* vol. 11, no. 18, p. 5450, 2022.

[7] A. Vephasayanant, A. Jitpattanakul, P. Muneesawang, K. Wongpatikaseree, and N. Hnoohom, "YOLO-based image segmentation for the diagnostic of spondylolisthesis from lumbar spine X-ray images," *IEEE Access,* vol. 12, pp. 182242-182258, 2024.

[8] C. Xu *et al.*, "Deep Learning-Based Diagnosis of Lumbar Spondylolisthesis Using X-Ray Imaging," *Diagnostics,* vol. 15, no. 16, p. 2015, 2025.

[9] F. N. M. Zamri, T. S. Gunawan, S. H. Yusoff, A. A. Alzahrani, A. Bramantoro, and M. Kartiwi, "Enhanced small drone detection using optimized YOLOv8 with attention mechanisms," *IEEE Access,* vol. 12, pp. 90629-90643, 2024.

[10] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132-7141.

[11] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3-19.

[12] Q. Guan, Y. Huang, Z. Zhong, Z. Zheng, L. Zheng, and Y. Yang, "Diagnose like a radiologist: Attention guided convolutional neural network for thorax disease classification," *arXiv preprint arXiv:1801.09927,* 2018.

[13] O. Oktay *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999,* 2018.

[14] G. Papanastasiou, N. Dikaios, J. Huang, C. Wang, and G. Yang, "Is attention all you need in medical image analysis? A review," *IEEE Journal of Biomedical and Health Informatics,* vol. 28, no. 3, pp. 1398-1411, 2023.

[15] P. Chen, Z. Zhou, H. Yu, K. Chen, and Y. Yang, "Computerized-Assisted Scoliosis Diagnosis Based on Faster R-CNN and ResNet for the Classification of Spine X-Ray Images," *Computational and Mathematical Methods in Medicine,* vol. 2022, no. 1, p. 3796202, 2022.

[16] Z. Al-Milaji and H. Yousif, "Lightweight deep learning model optimization for medical image analysis," *International Journal of Imaging Systems and Technology,* vol. 34, no. 5, p. e23173, 2024.

[17] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*, 2019: PMLR, pp. 6105-6114.

[18] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510-4520.

[19] A. Howard et al., "Searching for mobilenetv3," in Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 1314-1324.

[20] M. Tan *et al.*, "Mnasnet: Platform-aware neural architecture search for mobile," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 2820-2828.

[21] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700-4708.

[22] J. Irvin *et al.*, "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in *Proceedings of the AAAI conference on artificial intelligence*, 2019, vol. 33, no. 01, pp. 590-597.

[23] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding transfer learning for medical imaging," *Advances in neural information processing systems,* vol. 32, 2019.

[24] X. Liu *et al.*, "Advances in deep learning-based medical image analysis," *Health Data Science,* vol. 2021, p. 8786793, 2021.

[25] C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, "Key challenges for delivering clinical impact with artificial intelligence," *BMC Medicine,* vol. 17, no. 1, p. 195, 2019.

[26] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nature Medicine,* vol. 25, no. 1, pp. 44-56, 2019.

[27] S. Jetley, N. A. Lord, N. Lee, and P. H. Torr, "Learn to pay attention," *arXiv preprint arXiv:1804.02391,* 2018.

[28] H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu, and X. Cao, "Joint optic disc and cup segmentation based on multi-label deep network and polar transformation," *IEEE Transactions on Medical Imaging,* vol. 37, no. 7, pp. 1597-1605, 2018.

[29] V. V. Valindria *et al.*, "Multi-modal learning from unpaired images: Application to multi-organ segmentation in CT and MRI," in *2018 IEEE winter conference on applications of computer vision (WACV)*, 2018: IEEE, pp. 547-556.

[30] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested U-Net architecture for medical image segmentation," in *International workshop on deep learning in medical image analysis*, 2018: Springer, pp. 3-11.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[32] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data,* vol. 6, no. 1, pp. 1-48, 2019.

[33] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering,* vol. 21, no. 9, pp. 1263-1284, 2009.

[34] G. Douzas and F. Bacao, "Effective data generation for imbalanced learning using conditional generative adversarial networks," *Expert Systems with Applications,* vol. 91, pp. 464-471, 2018.

https://doi.org/10.31436/iiumej.v27i2.4361

TSGL-Rust: Teaching Concurrency Through Compile-Time Enforcement via Rust-Based Visualization of Classical Synchronization Problems

TEDDY SURYA GUNAWAN*

Electrical and Computer Engineering Dept., International Islamic University Malaysia, Malaysia

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

(Received: 10 March 2026; Accepted: 30 April 2026; Published online: 10 May 2026)

ABSTRACT: Concurrent programming remains a persistent challenge in undergraduate computer science and engineering education, largely because its failure modes are difficult to observe and reason about. Race conditions and deadlocks often emerge only under specific thread interleavings, leaving students to reconcile correct-looking code with unpredictable runtime behavior. Prior instructional approaches, such as the Thread-Safe Graphics Library (TSGL), address this issue by enabling real-time visualization, but they rely on languages that do not enforce safe access to shared state. This separation between observation and enforcement leaves a gap between how concurrency errors are detected and how they are prevented. This paper presents TSGL-Rust, a Rust-based reimplementation of classical TSGL visualizations that integrates compile-time safety guarantees with runtime visualization. Implemented using the egui/eframe framework, the system reproduces three canonical synchronization problems, namely Producer–Consumer, Reader–Writer, and Dining Philosophers, while shifting certain classes of errors from runtime to compilation through Rust's ownership model and borrow checker. The architecture deliberately separates concurrency logic from rendering through a crossbeam channel, decoupling worker threads from the egui main-thread render loop. This paper is presented as a system and tool description: it focuses on architectural design, the mapping from C++ TSGL primitives to Rust constructs, the channel-based communication pattern, and the instructional affordances that follow from these design choices. The contribution lies in the system itself and in the boundary it makes visible between data-race prevention and logical errors, such as deadlock, rather than in any claim about measured learning gains.

ABSTRAK: Pengaturcaraan serentak kekal sebagai cabaran utama dalam pendidikan sarjana muda sains komputer dan kejuruteraan, kerana mod kegagalannya sukar dicari dan difahami secara sistematik. Keadaan perlumbaan dan kebuntuan lazimnya hanya muncul di bawah jalinan pelaksanaan jalur tertentu, sekaligus menyebabkan pelajar perlu menyelaras kod yang kelihatan betul dengan tingkah laku masa jalan yang tidak menentu. Pendekatan pengajaran terdahulu, seperti Koleksi Grafik Jalur-Selamat (TSGL), menangani isu ini melalui visualisasi masa nyata, tetapi bergantung pada bahasa yang tidak menguatkuasa capaian selamat pada keadaan berkongsi. Pemisahan antara pemerhatian dan penguatkuasaan ini mewujudkan jurang antara bagaimana ralat serentak dikesan dan bagaimana ia dicegah. Kajian ini membentangkan TSGL-Rust, iaitu pelaksanaan semula TSGL berasaskan Rust yang mengintegrasi jaminan keselamatan masa kompilasi dan visualisasi masa jalan. Melalui rangka kerja e-GUI/e-rangka, sistem ini menjana semula tiga masalah menyegerakan kanonik, iaitu Pengeluar-Pengguna, Penulis-Pembaca, dan Dining Philosophers, sambil mengalihkan sebahagian kelas ralat daripada masa jalan kepada peringkat kompilasi melalui model pemilikan dan pemeriksa pinjaman Rust. Seni bina sistem ini secara sengaja memisahkan logik serentak daripada paparan alur silang, dengan jalur pekerja dipisahkan daripada gelung peparan utama eGUI. Kajian ini dibentangkan sebagai penerangan sistem dan alat: tumpuan

https://doi.org/10.31436/iiumej.v27i2.4361

diberi pada reka bentuk seni bina, pemetaan primitif TSGL C++ pada konstruk Rust, corak komunikasi berasaskan saluran, dan keupayaan instruksional yang berpunca daripada reka bentuk ini. Sumbangan utama terletak pada sistem ini sendiri dan pada sempadan yang dijelaskan antara pencegahan keadaan perlumbaan dan ralat logik seperti kebuntuan, dan bukan pada sebarang dakwaan terhadap pencapaian pembelajaran yang diukur.

KEYWORDS: *Concurrent programming, Rust, compile-time safety, visualization, operating systems education*

1. INTRODUCTION

Multicore processors now define the baseline computing environment, and concurrency has correspondingly become a required competency in modern computer science curricula [1]. Within operating systems and systems programming courses, students are expected to reason about interleavings, synchronization mechanisms, and subtle failure modes that arise under parallel execution. Despite this expectation, conceptual mastery often lags behind exposure. Concurrency resists intuition.

Among these difficulties, the invisibility of failure remains central. Race conditions emerge only under specific scheduling sequences, often fleeting and irreproducible. Deadlocks surface under particular resource acquisition patterns that may not appear during routine testing. Students frequently encounter programs that compile cleanly and appear correct, yet behave inconsistently at runtime. The resulting gap between static code and dynamic behavior complicates reasoning and may partly reflect the absence of structural feedback in conventional languages [2].

Prior work attempted to address this gap through visualization. The Thread-Safe Graphics Library (TSGL) was introduced in [3] to enable real-time rendering under multithreaded execution, allowing students to observe concurrency as it unfolds. Empirical results suggested improved comprehension of parallel constructs. Even so, TSGL operates within C++, where data races are undefined behavior with no compile-time guarantees. Visualization exposes symptoms but does not prevent their cause. Students still confront misbehavior after execution, often without clear guidance on its origin.

Rust introduces a different dynamic. Its ownership model and borrow checker enforce memory safety and eliminate data races at compile time [4]. Access to shared mutable state is restricted unless properly synchronized, and violations are rejected during compilation with precise diagnostics. The location of failure shifts. Instead of debugging unpredictable runtime outcomes, programmers encounter structured compiler feedback that links unsafe code patterns directly to concurrency violations. This shift narrows the conceptual distance between program structure and correctness, though its accessibility likely depends on prior familiarity with ownership semantics [5].

Building on this premise, TSGL-Rust reinterprets visualization through compile-time enforcement. Implemented in Rust using the egui immediate-mode GUI framework, the system reproduces three canonical synchronization problems, namely Producer-Consumer, Reader-Writer, and Dining Philosophers. Each visualization preserves the temporal dynamics emphasized in TSGL while embedding them within a language that prevents data races before execution. The tool was developed for an undergraduate Operating Systems course at IIUM, where Rust is a required component, and is presented here as a system and tool design rather than as an empirical educational study.

This work makes three contributions. First, a system architecture that integrates compile-time safety with runtime visualization for the three classical synchronization problems within a single Cargo workspace, separating concurrency logic from rendering via a crossbeam channel. Second, an explicit mapping from C++ TSGL primitives to their Rust equivalents. Third, an instructional affordance: compile-time diagnostics and runtime visualization together expose the boundary between data-race prevention and logical errors, such as deadlock—a boundary that is structural rather than empirical and follows directly from the choice of language and architecture.

2. Background and Related Works

Situating TSGL-Rust requires drawing together several strands of prior work: concurrency visualization tools, Rust's safety model, recent advances in Rust pedagogy, and dynamic analysis or model-checking systems for concurrency. The discussion below positions TSGL-Rust against each other.

2.1. Original TSGL

Adams et al. (2018) introduced the Thread-Safe Graphics Library (TSGL) [3], a C++11 framework that enables multiple threads to render to a shared canvas in near-real time. The design goal was pragmatic. Instructors could insert drawing calls into existing multithreaded programs with minimal restructuring, thereby making execution traces visible as graphical artifacts. Students no longer relied solely on textual output or static reasoning. They observed concurrency unfold. The original TSGL is available at https://github.com/Calvin-CS/TSGL/tree/visExpUpdates.

Empirical evaluation suggested measurable benefits. Students exposed to visualization demonstrated improved understanding of parallel constructs compared to those using traditional methods. More recent work in computing education has reinforced this general finding, indicating that dynamic visual representations can reduce cognitive load when reasoning about temporal processes [6, 7]. Despite this, TSGL inherits fundamental limitations from C++. Synchronization remains a matter of discipline rather than enforcement. A program containing a data race compiles successfully and may execute without immediate failure, leaving inconsistent behavior as the only evidence of the underlying fault. In practice, visualization reveals symptoms but offers little structural guidance toward correctness.

2.2. Rust's Safety and Concurrency Model

Rust approaches concurrency from a different premise. Memory safety and data race freedom are enforced through the ownership model and borrow checker, both evaluated at compile time [4]. A value is either uniquely owned, immutably borrowed by multiple references, or mutably borrowed by exactly one reference. These constraints are not advisory. The compiler rejects violations outright.

For concurrent programming, this design encodes synchronization into the type system itself. Shared ownership is implemented using constructs such as `Arc<T>`, while controlled mutation is mediated by `Mutex<T>` or `RwLock<T>`. Access to protected data is only possible through guard objects returned by lock acquisition, ensuring that unsynchronized access is structurally impossible. Research on Rust's formal semantics indicates that these guarantees extend even under relaxed memory models, providing strong safety properties without runtime overhead [5]. Evidence from large-scale systems further indicates that adopting memory-safe languages significantly reduces vulnerability density, particularly for concurrency-related

defects [8, 9]. Industry and government guidance have reinforced this position, with recent recommendations from the U.S. Office of the National Cyber Director advocating memory-safe languages for critical software [10].

These guarantees are bounded. Deadlock, starvation, and other liveness errors remain outside the scope of the type system. That boundary is pedagogically useful. It distinguishes clearly between what the language enforces and what must still be reasoned about explicitly.

2.3. egui and eframe

`egui` is an immediate-mode graphical user interface library for Rust that constructs the entire interface each frame from the current application state [11]. Unlike retained-mode frameworks, where UI elements persist across frames, `egui` treats the interface as a function of state, recalculated continuously. This design aligns closely with the requirements of concurrency visualization. State changes frequently. Visual feedback must follow immediately.

The accompanying `eframe` crate provides the application scaffold, including window management, event handling, and the render loop, which executes on the main thread. As a result, worker threads cannot directly manipulate the interface and must instead communicate state updates to the render loop. That constraint introduces a clear separation between concurrency logic and rendering. While initially restrictive, it may reduce incidental complexity by preventing shared mutable access to UI state across threads.

Immediate-mode rendering shifts the developer's mental model. Rather than updating widgets imperatively, the program describes what the interface should look like given the current state at each frame. `egui` also compiles to native binaries across major platforms and to WebAssembly, allowing deployment without external graphics dependencies. For undergraduate settings, this reduces setup overhead and broadens accessibility.

2.4. Recent Rust Pedagogy and Ownership Visualization

Beyond the language itself, recent work has examined how Rust is learned. Crichton et al. [12] developed a grounded conceptual model of Rust's ownership types, implemented as the Aquascope visualizer, and integrated it into a forked online edition of The Rust Programming Language. Their controlled study reported a measurable improvement in scores on an Ownership Inventory after exposure to permission-based visualizations of borrowing checking, relative to the baseline pedagogy. Subsequent work has extended these ideas into broader curricular guidance for systems courses delivered in Rust [13]. While these contributions target the static side of the language, i.e., the ownership and borrowing rules themselves, they share with TSGL-Rust an emphasis on visual representation as a means of access to otherwise opaque static analyses.

TSGL-Rust complements this body of work along a different axis. Rather than visualizing the borrow checker's internal reasoning, it accepts the compile-time guarantees as given and visualizes the runtime behavior of programs that survive them. The two perspectives are not in competition. A course that uses Aquascope-style visualizations for ownership during early lectures and TSGL-Rust during the synchronization unit would address both sides of the static–dynamic divide that Rust learners are reported to find difficult [12].

2.5. Visualization, Dynamic Analysis, and Model Checking

Several established systems analyze concurrent behavior, though their assumptions and instructional roles differ in important ways. Helgrind and ThreadSanitizer instrument program execution to detect data races dynamically. Java PathFinder explores execution paths through

model checking. TLA+ supports formal specification and verification of concurrent algorithms. These tools intervene before, during, or after execution. Violations are discovered retrospectively or through exhaustive search.

Table 1. Comparison of concurrency education and debugging tools.

Tool / System	Primary purpose	Timing of feedback	Type of feedback	Instructional role	Vis. + Safety
Helgrind / ThreadSanitizer	Dynamic data-race detection	Runtime	Warnings, stack traces	Post-hoc debugging; non-deterministic triggering	—
Java PathFinder / TLA+	Model checking and formal verification	Pre-execution (state exploration)	Counter-example traces	Formal-methods exposure; steep learning curve	—
Progvis [14-16]	Object-graph visualization with model checking	Runtime; verifier-supported	Object-graph rendering, feedback	Concurrency intuition with shared-memory focus	Vis.
TSGL [3]	Real-time concurrency visualization	Runtime only	Graphical thread-state rendering (C++ / OpenGL)	Conceptual insight into parallel behavior	Vis.
Aquascope / The Rust Book fork [12]	Ownership and borrow-checker visualization	Compile-time (static)	Permission diagrams alongside the source	Rust ownership pedagogy	Safety
TSGL-Rust (this work)	Visualization plus compile-time safety	Compile-time AND runtime	Compiler diagnostics + graphical thread-state rendering	Conceptual insight AND structural safety	Vis. + Safety

Note: "Vis." denotes real-time graphical visualization of program state. "Safety" denotes compile-time data-race prevention enforced by the type system. TSGL-Rust is the only tool in this comparison that provides both within a single environment.

Educational visualization operates under a different constraint. Rather than locating errors, it renders the temporal structure of concurrent execution visible as it unfolds. Progvis, developed by Strömbäck and colleagues [14], visualizes concurrent C and C++ programs as object graphs and has been extended with a weak memory model and integrated model checking [15] and with model-checking-driven feedback combined with peer grading [16]. These extensions move Progvis closer to the dynamic-analysis end of the spectrum while preserving its visualization-first character. TSGL [3] follows the visualization-first approach by exposing runtime behavior. TSGL-Rust extends this approach by introducing compile-time enforcement as an additional feedback channel. Table 1 contrasts these systems along dimensions relevant to undergraduate instruction.

TSGL-Rust occupies a hybrid position. It does not replace existing tools but shifts part of the feedback earlier: compile-time rejection prevents unsafe programs from executing, while runtime visualization remains essential for liveness errors such as deadlock and starvation, which lie beyond the reach of the type system. The two channels are complementary, i.e., one constrains what can be expressed; the other reveals how correct programs behave over time.

https://doi.org/10.31436/iiumej.v27i2.4361

Within this landscape, TSGL-Rust differs from Aquascope along the static–dynamic axis (static permissions versus runtime thread states), from Progvis along the language and safety axis (Rust with compile-time data-race prevention versus C/C++ with runtime model checking), and from the original TSGL along the enforcement axis (a memory-safe language versus an unsafe one). The novelty, therefore, lies less in any single capability than in the combination, and in the architectural choices that make this combination tractable for an undergraduate operating systems setting.

3. Classical Synchronization Problems

Classical synchronization problems provide structured entry points through which students can connect abstract concurrency concepts to concrete operating system behavior, while also exposing the boundary between compile-time enforcement and runtime correctness. Fig. 1 illustrates three classical synchronization problems, including Producer–Consumer, Reader–Writer, and Dining Philosophers. In practice, these problems arise in both process-based and thread-based concurrency, depending on how execution is scheduled and how memory is shared [17-19]. Processes may coordinate through pipes, sockets, or shared memory regions, while threads typically interact within a shared address space using synchronization primitives such as mutexes or condition variables. Despite these differences, the underlying challenge remains consistent. Multiple execution contexts must coordinate access to shared state without violating correctness.

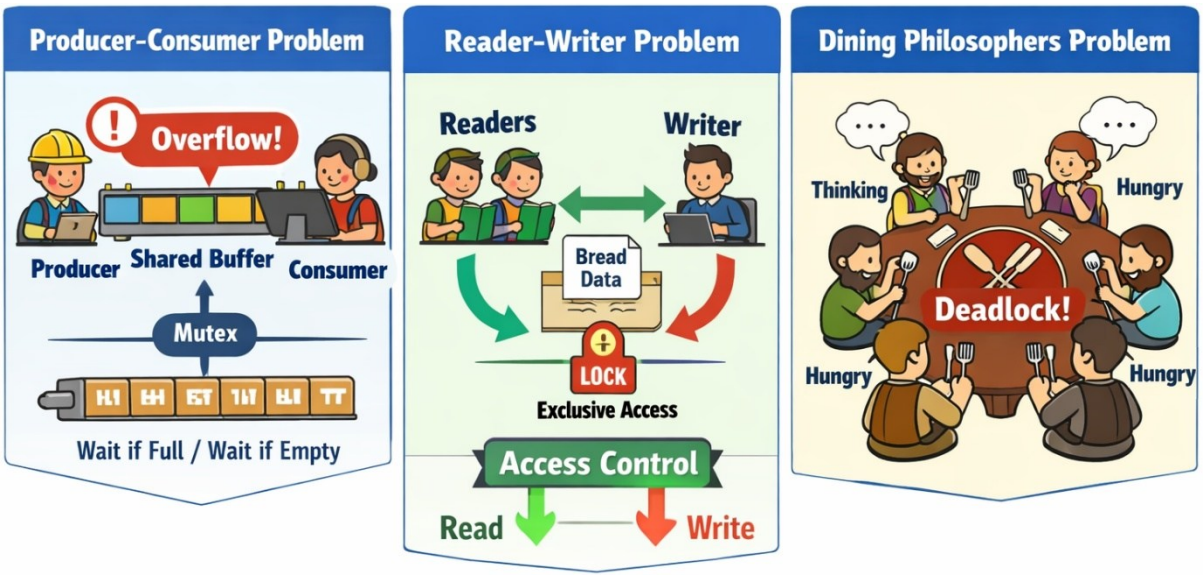


Figure 1. Classical synchronization problems addressed by TSGL-Rust: Producer–Consumer, Reader–Writer, and Dining Philosophers.

Resources involved in these interactions can be categorized as consumable or reusable, and this distinction shapes how synchronization is designed. Consumable resources are produced and then removed once used, as in a Linux pipe or Windows message queue, where data flows from producer to consumer and does not persist. Reusable resources, by contrast, can be acquired and released repeatedly, such as a mutex protecting a kernel structure or a file lock shared among processes. Consumable resources emphasize availability and ordering, while reusable resources emphasize controlled access and ownership. These patterns appear frequently in operating systems [20-22], though they may not be immediately visible to system users. Grounding the classical problems in such familiar mechanisms makes the abstractions

feel less artificial, and the role of synchronization clearer as a recurring constraint in system design rather than an isolated curricular topic.

3.1. Producer–Consumer Problem

The Producer–Consumer problem models coordination over a bounded shared buffer between producers that generate data and consumers that process it. Correct behavior depends on mutual exclusion for buffer access and condition-based blocking when the buffer becomes full or empty. Under incorrect synchronization, subtle failures emerge. Data may be overwritten, duplicated, or lost entirely. These errors often appear intermittently, which makes them particularly difficult to diagnose.

In practice, this pattern appears throughout operating systems. A Linux shell pipeline such as `ls | grep` implements a bounded pipe buffer managed by the kernel. In Windows, anonymous pipes and producer–consumer queues in the Win32 API follow the same structure, with blocking semantics enforced by the system. Print spoolers in both environments exhibit identical behavior: applications produce jobs, and a background service consumes them. The abstract buffer is not merely theoretical. It exists in tools that are routinely used. Table 2 summarizes representative operating system instances of this pattern.

Table 2. Real Operating System Examples of the Producer-Consumer Pattern.

Real-World Instance	Producer	Consumer	Buffer
Linux pipe (`ls | grep`)	ls process (`stdout`)	grep process (`stdin`)	OS pipe buffer
Windows anonymous pipe	Writing process/thread	Reading process/thread	OS-managed pipe buffer
Print spooler	User applications	Print service/daemon	Spool queue (disk/memory)
Network socket	Application (`send()`)	OS network stack	Socket send buffer
Video streaming	Decoder thread	Renderer thread	Frame ring buffer

The shell pipeline is the most accessible example. It is routinely used without explicit recognition that blocking and synchronization are being enforced beneath the surface.

3.2. Reader–Writer Problem

The Reader–Writer problem addresses controlled access to shared data that can be read concurrently but must be written exclusively. Multiple readers may proceed in parallel. Writers require exclusive access. The central challenge lies in balancing fairness and throughput, since prioritizing one class of operations may delay the other.

This pattern aligns closely with language-level abstractions. Rust's `RwLock<T>` encodes the distinction between shared and exclusive access directly in the type system. A thread may hold multiple read guards or a single write guard, but never both simultaneously. This constraint is enforced at compile time, shaping how access must be structured before execution begins. It shifts part of the reasoning earlier, though understanding the implications of fairness policies still requires runtime observation.

Operating systems rely heavily on this pattern. Linux employs reader–writer locks (`rwlock_t`) in subsystems such as virtual memory management. Windows provides Slim Reader/Writer (SRW) locks for efficient shared access. File metadata, DNS caches, and server configuration structures all depend on this form of synchronization. Table 3 summarizes representative instances.

https://doi.org/10.31436/iiumej.v27i2.4361

Table 3. Real Operating System Examples of the Reader-Writer Pattern.

Real-World Instance	Readers	Writers	Shared Resource
Linux `rwlock_t`	CPU cores (page reads)	Core updating page table	Virtual memory page table
Windows SRW lock	Multiple threads (read access)	Single thread (write access)	Shared memory region
File system inode	Processes calling `stat()`	Process calling `chmod()`	`Inode` metadata
DNS resolver cache	DNS lookup threads	Cache refresh process	Cached DNS entries
Web server configuration	Worker threads (HTTP requests)	Admin reload operation	Configuration structure

This problem often appears straightforward until contention increases. Under load, the interaction between readers and writers becomes visible, and fairness policies take on greater importance.

3.3. Dining Philosophers Problem

The Dining Philosophers problem models resource contention in the presence of circular dependencies. Each participant requires multiple shared resources, and if all attempt to acquire them in the same order, a circular wait can arise. No participant can proceed. The system halts.

Unlike the previous problems, this one exposes a limitation of compile-time guarantees. Proper use of mutexes ensures memory safety, yet deadlock remains possible. The failure arises not from unsafe access but from the ordering of resource acquisition. This distinction often becomes clearer through observation than through abstract reasoning alone.

Operating system examples are widespread. In Linux, kernel subsystems acquiring multiple mutexes may deadlock if lock ordering is inconsistent, prompting the use of runtime tools such as lock dependency checkers. In Windows, file and handle locking across threads can produce similar circular waits. Database systems provide a particularly concrete example. Transactions acquiring row-level locks may form a cycle, leading to deadlock detection and forced rollback. Table 4 summarizes representative instances.

Table 4. Real Operating System Examples of the Dining Philosophers (Deadlock) Pattern.

Real-World Instance	Actors	Resources	Deadlock Scenario
Database transactions	Concurrent transactions	Row-level locks	Circular wait detected (e.g., transaction rollback)
Multi-threaded file ops	Processing threads	File locks	Thread A waits for file2, Thread B waits for file1
Linux kernel mutexes	Kernel subsystems	Kernel mutexes	Lock ordering violation detected at runtime
Windows handle locking	Threads/processes	File/handle locks	Circular dependency across resource acquisition
GPU shader execution	Shader threads	Memory banks	Cyclic resource dependency

Database deadlocks are often the most visible. Errors such as transaction rollback occur in practice, yet the underlying pattern is rarely recognized without explicit instruction.

3.4. Comparative Synthesis and Rust Correctness Boundaries

Table 5 summarizes the three problems with respect to their primary hazard, synchronization mechanism, corresponding Rust abstraction, representative operating system example, and the extent to which Rust prevents the associated error.

Table 5. Comparative Summary of the Three Classical Synchronization Problems.

Property	**Producer–Consumer**	**Reader–Writer**	**Dining Philosophers**
Primary hazard	Buffer race	Read–write conflict	Deadlock/starvation
Key primitive	Mutex + Condvar	RwLock	Mutex per resource
Rust abstraction	`Arc<Mutex<VecDeque>>`	`Arc<RwLock<T>>`	`Vec<Arc<Mutex<()>>>`
OS example	Linux/Windows pipe	Kernel/SRW locks	DB/file locking
Prevented by Rust	Yes (data race)	Yes (data race)	No (logical error)

The final row carries particular weight. Rust eliminates data races in the first two problems through compile-time enforcement. Deadlock remains. The result is a layered model of correctness, in which some classes of errors are prevented structurally while others must still be reasoned about through execution and design.

3.5. Extension to Deep Learning Training and Inference Workflows

The correctness boundaries identified in Section 3.4 are not specific to operating systems. They recur at the application level in modern artificial intelligence workloads, where they govern the throughput and correctness of training and inference systems. The Producer–Consumer pattern appears most directly in the data loading pipeline. PyTorch's DataLoader with `num_workers > 0` spawns worker processes that produce preprocessed and augmented mini-batches into a bounded queue, while the training loop on the GPU consumes them. The `prefetch_factor` parameter sets the buffer capacity, and the same starvation and saturation conditions illustrated in Section 3.1 govern whether the GPU stalls or the CPU workers block. Frame ingestion pipelines for video tasks, i.e., capture, preprocessing, inference, rendering, chain several producer–consumer stages with bounded buffers between them. Asynchronous logging of training metrics and checkpoint writes follows the same pattern, but at a lower volume.

The Reader–Writer pattern surfaces in inference serving. Many concurrent request handlers read the model weights to perform forward passes, while a control thread occasionally swaps the weights when a new checkpoint is hot-loaded. Shared lookup tables for physics priors, for example, precomputed turbulence kernels or atmospheric coefficients indexed by parameters such as the refractive index structure constant, exhibit the same readers-many, writer-rare structure. The Dining Philosophers pattern is less direct in this domain but can arise in distributed training, where lock-ordering errors across gradient synchronization primitives can cause hangs in multi-GPU settings. Within this broader landscape, the structural distinction that TSGL-Rust makes visible, between data-race prevention and liveness, has currency beyond the systems curriculum in which it is currently deployed, and the same enforcement boundary that separates Sections 3.1 and 3.2 from 3.3 reappears in AI pipelines that combine high-throughput data movement with shared mutable state.

4. SYSTEM ARCHITECTURE

This section describes the architecture of TSGL-Rust, covering the workspace organization, the mapping of C++ concurrency primitives to equivalent Rust constructs, and the channel-

https://doi.org/10.31436/iiumej.v27i2.4361

based communication pattern that separates worker thread logic from the egui rendering pipeline.

4.1. Design Overview

TSGL-Rust is organized as a Cargo workspace containing three independent binary targets: `reader_writer`, `producer_consumer`, and `dining_philosophers`. Each binary is self-contained, with its own concurrency logic and visualization code. The workspace uses a shared `Cargo.toml` declaring common dependencies: `eframe` and `egui` for visualization, `parking_lot` for efficient mutex and `RwLock` implementations, and `crossbeam` for multi-producer, multi-consumer channels. The overall system architecture is illustrated in Figure 2.

The central architectural decision is the separation of concurrency logic from rendering. In the original TSGL, drawing calls are made directly from worker threads into the shared Canvas, which is internally synchronized. In TSGL-Rust, worker threads never touch the GUI. Instead, each thread computes its state and transmits it to the main thread via a `crossbeam` channel. The `egui` render loop, running on the main thread at each frame, reads the latest state snapshot and renders the visualization. This design is required by `egui`'s threading model, but it also yields a cleaner separation of concerns: unit tests can verify synchronization behavior without instantiating any GUI, and changes to the visualization style require no changes to thread code, i.e., a structural guarantee that TSGL's shared-canvas approach cannot provide.

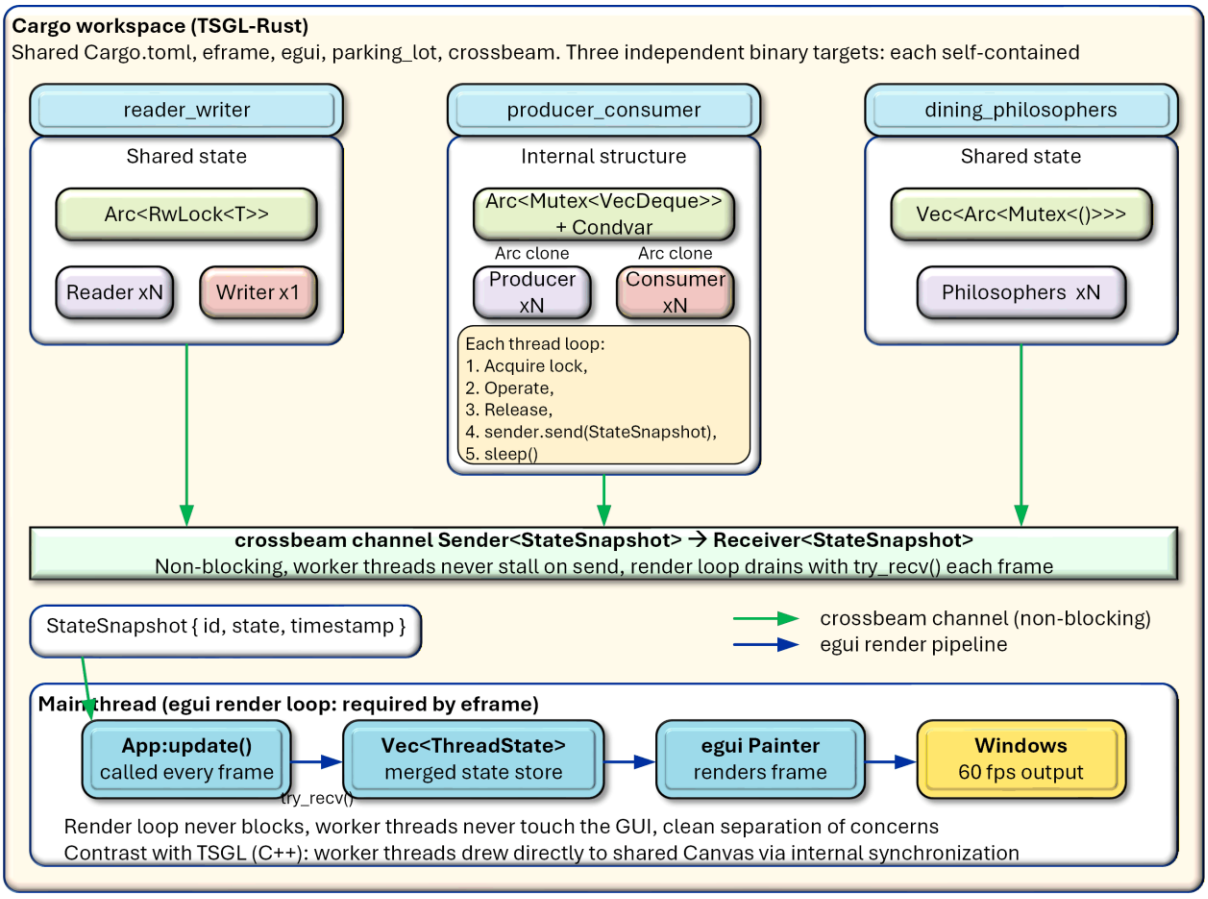


Figure 2. Overall system architecture of TSGL-Rust illustrating the Cargo workspace, worker thread concurrency logic, crossbeam channel, and egui render loop separation

4.2. C++ to Rust Primitive Mapping

Table 6 shows the direct mapping between C++ concurrency primitives used in TSGL and the corresponding Rust constructs used in TSGL-Rust. The most consequential difference is not syntactic but semantic. In C++, `std::mutex` provides mutual exclusion, but nothing in the type system prevents a thread from accessing the protected data without acquiring the lock. In Rust, `Arc<Mutex<T>>` wraps the data inside the lock: the protected value is only accessible through the lock guard returned by `lock()`, and the compiler rejects any attempt to access it otherwise. The synchronization requirement is encoded in the type, not enforced by programmer discipline. This distinction applies equally to `Arc<RwLock<T>>` in the `reader_writer` binary and to `Vec<Arc<Mutex<()>>>` in `dining_philosophers`: in each case, the relationship between the synchronization primitive and the data it protects is structurally enforced at compile time rather than assumed at runtime.

Table 6. Mapping of C++ TSGL concurrency primitives to Rust TSGL-Rust equivalents

C++ (TSGL)	Rust (TSGL-Rust)	Notes
`std::mutex`	`Arc<Mutex<T>>`	Compile-time lock enforcement
`std::condition_variable`	`Arc<Condvar>`	Paired with `Mutex`
`std::shared_mutex`	`Arc<RwLock<T>>`	`ReaderWriter` only
`std::thread`	`std::thread::spawn`	Returns `JoinHandle`
`shared_ptr<T>`	`Arc<T>`	Atomic reference counting
TSGL Canvas draw calls	egui Painter + channels	State sent to render thread
`std::this_thread::sleep_for`	`std::thread::sleep`	Direct equivalent

4.3. Thread Communication Pattern

All three programs follow a common communication pattern. Each worker thread is spawned with an `Arc` clone of a `crossbeam` `Sender`, moved into the thread closure at spawn time. At each state transition, a consumer removes an item; a philosopher changes from thinking to hungry; a writer acquires the database; the thread constructs a lightweight `StateSnapshot` struct containing the thread ID, current logical state, and a timestamp, and transmits it via `try_send()`. Using `try_send` rather than the blocking send ensures that worker threads are never stalled by rendering latency. If the channel buffer is momentarily full, the snapshot is silently discarded, and the thread proceeds immediately. Under normal operating conditions, the render loop drains the channel faster than threads produce updates, so snapshot loss is rare and inconsequential for an educational visualization.

The render loop calls `try_recv()` each frame to drain all available updates, merging each snapshot into a `Vec<ThreadState>` indexed by thread ID. This `Vec` always reflects the most recent known state of every thread, regardless of how many snapshots were transmitted since the last frame, and it is this `Vec` that the egui `Painter` renders at each update cycle. The pattern guarantees that the render loop never blocks waiting for thread updates and that the concurrency logic is never delayed by the rendering pipeline: a clean, deterministic boundary that reflects the ownership discipline Rust enforces throughout the codebase.

5. IMPLEMENTATION

Each of the three TSGL-Rust binaries follows the same channel-based architecture described in Section 4, but differs in the shared state type, synchronization primitives, and visual layout that are appropriate to its specific concurrency problem.

5.1. Reader-Writer

The Reader-Writer implementation uses `Arc<RwLock<i32>>` as the shared database. Multiple reader threads call `read()` on the `RwLock` in a loop, holding the read guard for a randomized duration representing reading time, then sleeping before the next read. A single writer thread calls `write()`, holds the exclusive guard for a shorter duration, and increments the protected value. Figure 3 shows the internal block structure of this binary.

The key insight is the compile-time box in Fig. 3. In Rust, a thread holding a read guard cannot call any method that mutates the protected value, in which the type system makes mutation via `&RwLockReadGuard` a compile-time error rather than a runtime failure. First-reader preference emerges naturally from `parking_lot`'s `RwLock` implementation with no explicit counter management by the programmer.

The visualization renders a left panel showing active reader count and a right panel showing writer state. A central rectangle representing the shared database pulses when a write is in progress. Thread state badges, including Reading, Waiting, and Writing, use blue for active, orange for waiting, and green for completed states. A speed control slider adjusts sleep durations, allowing the simulation to be slowed to a rate suitable for classroom demonstration.

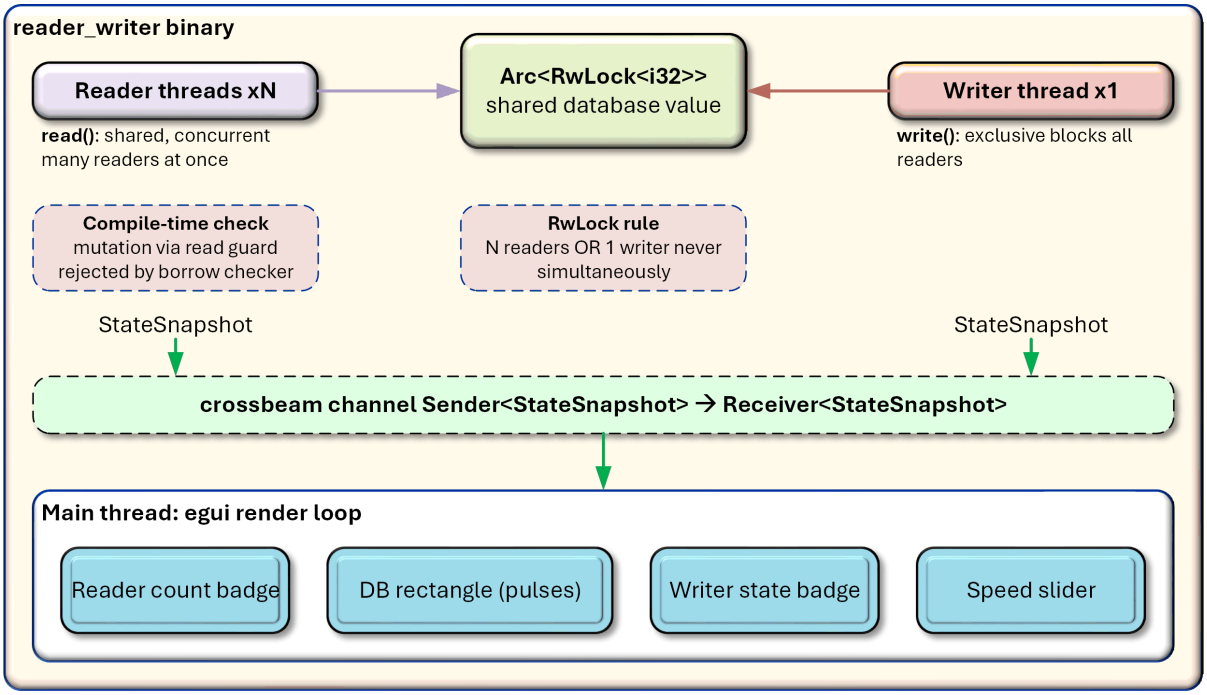


Figure 3. Reader-Writer Internal Block Structure

5.2. Producer-Consumer

The Producer-Consumer implementation uses `Arc<Mutex<VecDeque<u32>>>` as the bounded buffer, with a capacity constant configurable at compile time. `Condvar` pairs are used for blocking: producers call `wait()` on a full condition variable when the buffer reaches capacity, and consumers call `wait()` on an empty condition variable when the buffer is empty. Figure 4 shows the internal block structure of this binary.

The `Condvar::wait()` semantics box in Fig. 4 is worth noting. The method atomically releases the Mutex, suspends the calling thread, and then re-acquires the lock when signaled. This prevents the missed-wakeup bug that arises when a thread checks the condition and then releases the lock as separate operations, which is a common error in manual C implementations.

In Rust, `Condvar::wait()` requires a `MutexGuard` argument, making the lock relationship explicit in the type signature and matching the semantics of `std::condition_variable::wait()` in C++ while preventing access to the protected data without the guard.

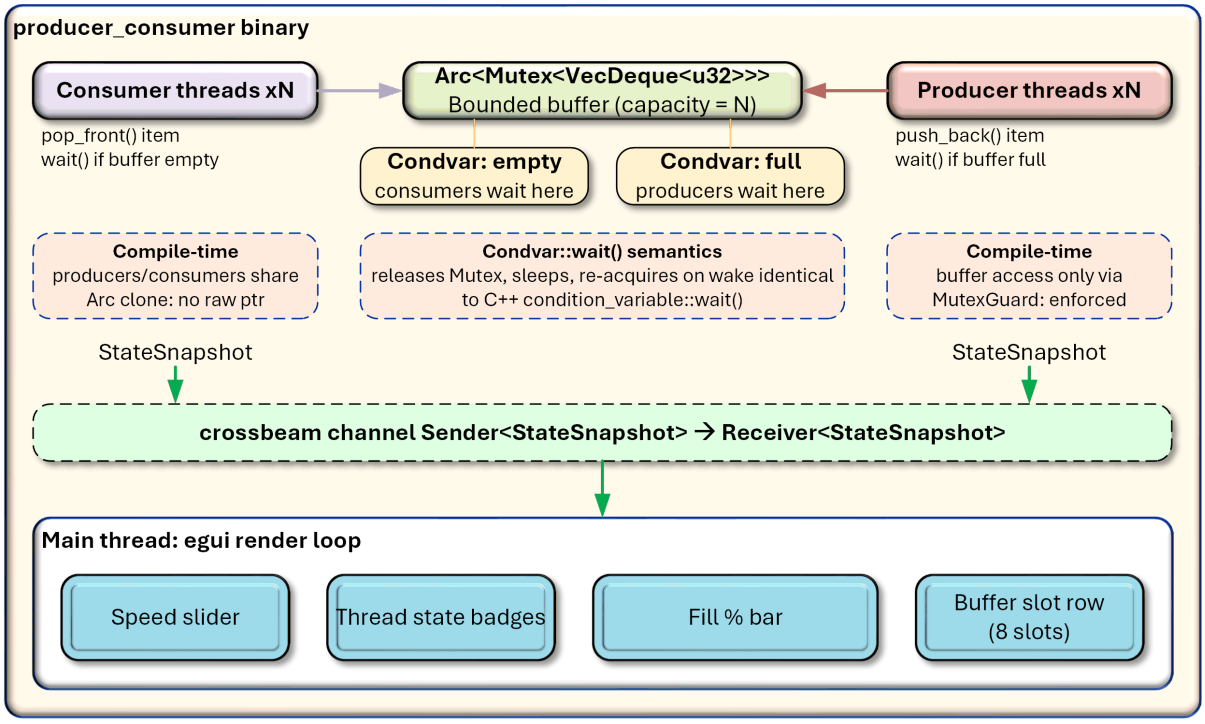


Figure 4. Producer-Consumer Internal Block Structure

The visualization renders a horizontal row of buffer slots, each colored blue when occupied and light gray when empty, with a fill percentage bar above. Producer threads appear on the left with directional indicators pointing into the buffer; consumer threads appear on the right pulling from it. When all slots are gray, consumers block; when all slots are blue, producers block. Saturation and starvation conditions are immediately visible without reading any code.

5.3. Dining Philosophers

The Dining Philosophers implementation uses a `Vec<Arc<Mutex<()>>>` with one mutex per fork, following the resource ordering deadlock avoidance strategy: each philosopher acquires the lower-numbered fork before the higher-numbered fork, breaking the circular wait condition identified by Dijkstra [17, 18]. Figure 5 shows the internal block structure of this binary.

The red dashed box in Fig. 5 is the most important teaching point of the entire section. Rust's borrow checker prevents data races, i.e., concurrent unsynchronised access to shared memory, but it cannot prevent deadlock, because deadlock is a logical error arising from lock acquisition order rather than from unsafe memory access. If the resource ordering strategy is removed and all philosophers acquire forks in the same ascending order, the program compiles without error and deadlocks at runtime. This behavior is intentional in TSGL-Rust. The Dining Philosophers visualization illustrates that Rust's compile-time guarantees are powerful yet bounded, and that correct synchronization logic remains the programmer's responsibility.

The visualization renders a circular table using the `egui::Painter` API with philosopher nodes distributed evenly around the circumference, coloured grey for thinking, orange for hungry, and green for eating. Fork segments between adjacent nodes are coloured to indicate availability. The circular layout makes the potential for circular waiting geometrically apparent, allowing us to trace the dependency ring that would form if all philosophers simultaneously held their left fork and waited for their right to be traced visually.

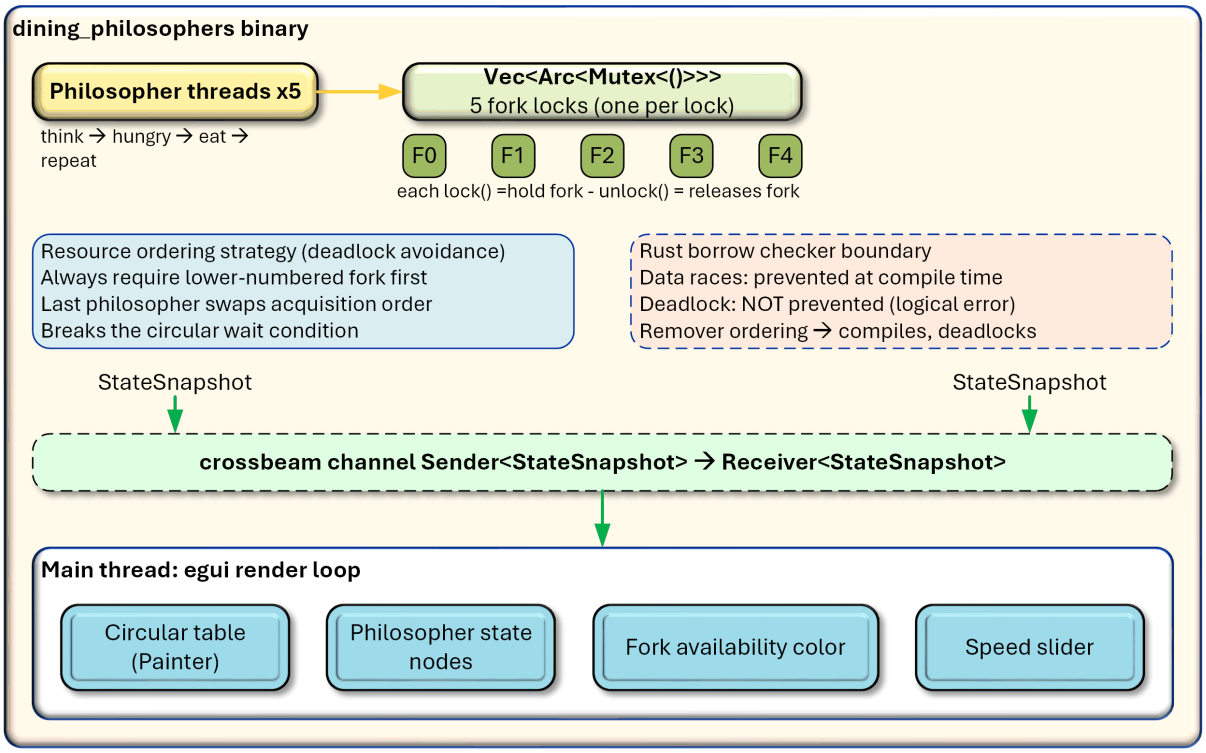


Figure 5. Dining-Philosophers Internal Block Structure

The complete source code for TSGL-Rust, including all three binary targets (`reader_writer`, `producer_consumer`, and `dining_philosophers`), the shared `Cargo.toml` workspace configuration, and the `README.md` project documentation, is publicly available at https://github.com/tsgunawan/TSGL_rust. The repository is structured to allow immediate execution on Windows, Linux, and macOS via a single `cargo run --bin <target>` command without any external graphics library installation, lowering the barrier to adoption for instructors who wish to integrate the tool into an existing operating systems laboratory session. Each binary is accompanied by inline code comments that map Rust synchronization constructs to their C++ equivalents, providing a side-by-side reference to support students transitioning from a C++-based concurrency curriculum. Instructors can fork the repository and modify the thread count, buffer capacity, and sleep duration constants, all of which are declared as named compile-time constants at the top of each source file, to produce variants suitable for specific classroom demonstrations or laboratory assessments.

6. Comparative Analysis

Comparing TSGL and TSGL-Rust reveals how compile-time enforcement and runtime visualization interact to shape both correctness guarantees and the architectural separation between concurrency logic and rendering.

6.1. Safety Guarantees

Table 7 compares the two systems with respect to core concurrency safety properties. The first row carries the most immediate pedagogical consequence. In C++ TSGL, a data race is undefined behavior. A program may compile, execute, and produce incorrect results without any warning. The presence of a race must be inferred from inconsistent output. In Rust TSGL-Rust, the same unsafe access pattern is rejected at compile time. The compiler reports the exact location and nature of the violation, requiring the structural cause to be addressed before execution begins.

Table 7. Safety property comparison between C++ TSGL and Rust TSGL-Rust.

Property	**C++ TSGL**	**Rust TSGL-Rust**
Data race prevention	Runtime (undefined behavior)	Compile-time (borrow checker)
Deadlock prevention	None	None (logical error)
Memory leak prevention	Manual / `shared_ptr`	Automatic (ownership)
Null pointer safety	Runtime crash	Compile-time (`Option<T>`)
Thread safety of canvas	Library-enforced (TSGL)	Type-enforced (`Arc`)

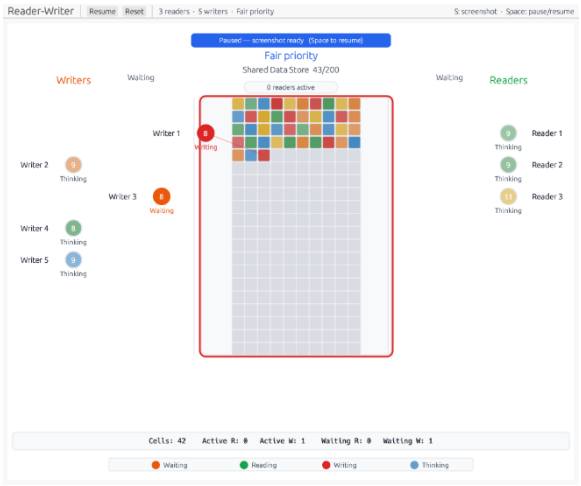
Not all classes of errors are eliminated. Deadlock remains possible in both systems. This separation matters because it distinguishes memory safety from logical correctness. Some errors are prevented by the language; others persist as design problems that require reasoning about execution order.

6.2. Visualization and Portability

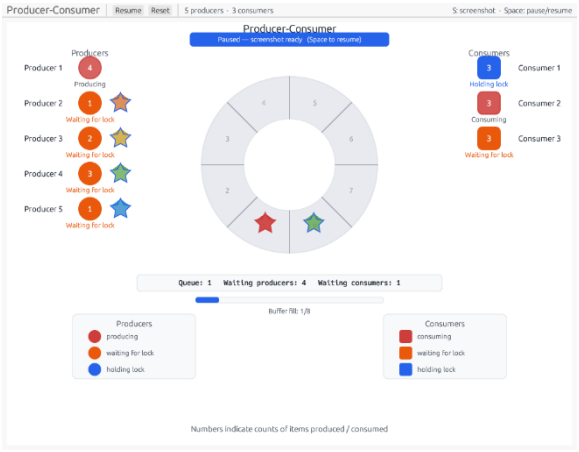
Both TSGL and TSGL-Rust provide real-time visualization of thread states, yet they differ in how easily those visualizations can be deployed and extended. TSGL relies on a custom OpenGL-based canvas, which introduces non-trivial installation requirements, particularly on Windows systems where additional graphics dependencies must be configured manually. TSGL-Rust, by contrast, builds using Cargo and the eframe framework without requiring external graphics libraries. Cloning the repository and executing a binary with a single command reduces setup overhead in laboratory environments.

Fig. 6 presents representative screenshots of the three classical synchronization problems implemented in TSGL-Rust. The visualizations expose thread states, resource availability, and blocking conditions directly, removing the need for textual debugging output to follow execution. The immediate-mode rendering model of `egui` ensures that each frame reflects the current system state without retaining outdated UI elements, which produces a direct mapping between program state and visual output.

Portability extends beyond installation. `egui` supports compilation to WebAssembly, enabling browser-based execution. A user without a local Rust toolchain could interact with the visualizations through a web interface. Such deployment is not available in TSGL's OpenGL-based design. While this capability is not yet fully exercised in the current implementation, it suggests a pathway toward broader accessibility in future iterations.



(a) reader_writer 3 5



(b) producer_consumer 5 3



(c) dining_philosophers 5

Figure 6. Sample Screenshots of TSGL-Rust: (a) Reader–Writer with 3 readers and 5 writers under fair priority. (b) Producer–Consumer with 5 producers and 3 consumers sharing the bounded buffer. (c) Dining Philosophers with 5 philosophers under the resource-ordering deadlock-avoidance strategy.

6.3. Architectural Complementarity

TSGL-Rust is not intended to replace TSGL. The two systems emphasize different points in the development cycle. TSGL exposes runtime behavior, allowing incorrect synchronization to be observed during execution. TSGL-Rust shifts part of that experience earlier by preventing certain classes of errors before the program runs.

Used in sequence, the two systems illustrate the same problem from complementary angles. A race condition that appears in C++ TSGL only as inconsistent output is rejected at compile time in Rust, with diagnostics that identify the unsafe access pattern. One system shows what goes wrong; the other constrains what can be written. Yet the combination highlights a limit rather than removing it: deadlock remains visible only at runtime in both systems, since liveness errors lie beyond the scope of static analysis. The resulting instructional affordance is a single environment in which the boundary between data-race prevention and liveness is made structurally explicit; whether and how instructors exercise that affordance lies outside the scope of the present system description.

7. DISCUSSION

This section discusses the design constraints that follow from the choices documented above, the practical context of the IIUM deployment, and the limitations of the current system.

7.1. Design Constraints and Trade-offs

Several design constraints emerge from situating TSGL-Rust within an undergraduate operating systems setting. Rust's ownership model introduces a conceptual overhead that must be in place before concurrency primitives can be used productively. When ownership and synchronization are introduced in close succession, the resulting cognitive load is substantial [23]. Borrow checker diagnostics, while precise, can initially appear opaque without prior grounding in the underlying permissions model [12]. This is not a property of TSGL-Rust as such, but a constraint inherited from the language; it shapes how and when the tool can be introduced in a course.

Deadlock remains unaffected by Rust's guarantees. The Dining Philosophers implementation makes this boundary explicit. Programs that satisfy all compile-time constraints may still fail at runtime due to incorrect lock acquisition order. The separation between memory safety and liveness properties is intrinsic to the language design, and TSGL-Rust surfaces it rather than resolves it.

On the rendering side, the immediate-mode model of egui required a redesign rather than a direct translation of TSGL visualizations. The resulting interface presents thread states and resource availability through a different visual idiom than the original OpenGL canvas. This is a design consequence of the framework choice, traded against the elimination of external graphics dependencies and the resulting portability gains.

7.2. Deployment Context

TSGL-Rust was developed for and deployed in the EECE 4342 Operating Systems course at IIUM, in which Rust proficiency is a stated outcome. Within that course, the tool is positioned after an ownership-focused introduction to Rust and is used during the synchronization unit alongside conventional lecture material on the three classical problems. The placement reflects the conceptual prerequisite identified above: that exposure to ownership semantics in isolation makes subsequent borrow checker diagnostics in concurrent code more interpretable. This sequencing is described here as a deployment choice rather than as a tested intervention; no controlled comparison was conducted.

Table 8. Comparative Characteristics of TSGL and TSGL-Rust

Dimension	**TSGL (C++)**	**TSGL-Rust**
Error detection	Runtime only	Compile-time and runtime
Feedback form	Indirect (output discrepancy)	Direct (compiler diagnostic)
Determinism of rejection	Low (race-dependent)	Higher (compile-time rejection)
Reasoning supported	Temporal	Structural and temporal

The Cargo-based build, the absence of external graphics dependencies, and the single-command execution model were all driven by deployment pragmatics in this setting. Laboratory machines run a mix of Windows, Linux, and macOS, and prior experience with TSGL on Windows highlighted the cost of platform-specific OpenGL setup. Removing that cost was a primary motivation for choosing egui/eframe.

Table 8 summarizes the comparative characteristics of TSGL and TSGL-Rust along dimensions that follow directly from these design and deployment choices. The dimensions are technical: error-detection timing, the form of feedback, the determinism of rejection, and the type of reasoning the tool supports. Claims about the effect of these characteristics on student learning lie outside the scope of this paper and would require a separate, controlled study to establish.

7.3. Limitations

The contribution of this paper is a system and a design. Several limitations follow from that scope and should be stated plainly. No controlled experiment was conducted. No quantitative learning outcomes were measured. Comparisons with TSGL are technical and structural, not empirical, and any references to instructional context describe deployment circumstances rather than tested effects.

The visualization style differs from the original TSGL due to a change in rendering framework, and a side-by-side perceptual comparison was not conducted. The WebAssembly deployment path is supported by the egui toolchain in principle, but has not been exercised end-to-end in the current implementation. Coverage is restricted to the three canonical problems addressed; other classical problems, such as the Sleeping Barber or the Cigarette Smokers, would require additional binaries following the same architectural pattern.

Finally, the boundary that TSGL-Rust makes visible, i.e., the line between data-race prevention and liveness, is a property of Rust and of the chosen synchronization primitives. The system surfaces this boundary; it does not extend it. Tools such as Progvis [14], which integrate model checking, sit on the other side of that line and would complement TSGL-Rust in a curriculum that aims to address liveness systematically. Combining the two would be an interesting direction for future work, but it is not attempted here.

8. CONCLUSIONS

TSGL-Rust is a Rust-based reimplementation of the Producer–Consumer, Reader–Writer, and Dining Philosophers visualizations. It preserves the temporal visibility of the original TSGL and combines it with compile-time data-race prevention, yielding a single environment in which compiler diagnostics and runtime visualization operate side by side.

The technical contribution is threefold. First, an architecture that decouples concurrency logic from rendering through a channel between worker threads and the immediate-mode render loop, i.e., a separation that the original shared-canvas design cannot provide. Second, an explicit mapping from C++ synchronization primitives to their Rust equivalents, with the synchronization requirement encoded in the type rather than enforced by discipline. Third, a build path that requires no external graphics dependencies, enabling consistent execution across Windows, Linux, and macOS without platform-specific OpenGL setup.

From these properties follows an instructional affordance rather than a measured effect. By rejecting unsafe access patterns at compile time and visualizing the runtime behavior of programs that survive the borrow checker, TSGL-Rust makes the boundary between data-race prevention and liveness structurally visible, with deadlock deliberately exposed in the Dining Philosophers binary as a logical error beyond the reach of the type system.

Whether this affordance translates into learning outcomes is an empirical question this paper does not answer. A controlled study comparing comprehension and debugging behavior under matched conditions is the principal direction for future work, alongside end-to-end

WebAssembly deployment, integration with model-checking tools that address liveness, and extension toward concurrency primitives in deep learning training and inference pipelines, where the same producer–consumer and reader–writer structures recur at a different layer of the stack.

ACKNOWLEDGMENT

This work is partially supported by and also conceptually aligned with ongoing research under the AOARD project FA2386-25-1-4061 and SPI25-295-0295, "Adaptive Physics-Guided Transformer with Explainable Attention for Atmospheric Turbulence Mitigation," particularly in relation to concurrent data acquisition, inference, and visualization pipelines. The author thanks the students of the EECE 4342 Operating Systems course at IIUM for their engagement during the development and classroom deployment of TSGL-Rust. The author used generative AI tools (ChatGPT and Claude) solely for language refinement and structural editing during manuscript preparation; all AI-assisted text was reviewed and verified by the authors, and no AI was used to generate research data, results, or substantive technical content.

REFERENCES

[1] A. N. Kumar *et al.*, "Computer science curricula 2023," ed: ACM, 2024.
[2] C. Sadowski, E. Söderberg, L. Church, M. Sipko, and A. Bacchelli, "Modern code review: A case study at Google," in *Proceedings of the 40th International Conference on Software Engineering: Software Engineering in Practice*, 2018, pp. 181-190.
[3] J. C. Adams *et al.*, "TSGL: A tool for visualizing multithreaded behavior," *Journal of Parallel and Distributed Computing,* vol. 118, pp. 233-246, 2018.
[4] S. Klabnik and C. Nichols, *The Rust programming language*. No Starch Press, 2023.
[5] H.-H. Dang, J.-H. Jourdan, J.-O. Kaiser, and D. Dreyer, "RustBelt meets relaxed memory," *Proceedings of the ACM on Programming Languages,* vol. 4, no. POPL, pp. 1-29, 2019.
[6] J. C. Castro-Alonso, P. Ayres, and J. Sweller, "Instructional visualizations, cognitive load theory, and visuospatial processing," in *Visuospatial processing for education in health and natural sciences*: Springer, 2019, pp. 111-143.
[7] R. Duran, A. Zavgorodniaia, and J. Sorva, "Cognitive load theory in computing education research: A review," *ACM Transactions on Computing Education (TOCE),* vol. 22, no. 4, pp. 1-27, 2022.
[8] R. Jung, J.-H. Jourdan, R. Krebbers, and D. Dreyer, "Safe systems programming in Rust," *Communications of the ACM,* vol. 64, no. 4, pp. 144-152, 2021.
[9] S. Panter and N. Eisty, "Rusty Linux: Advances in Rust for Linux kernel development," in *Proceedings of the 18th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, 2024, pp. 496-502.
[10] W. House, "Back to the building blocks: A path toward secure and measurable software," Office of the National Cyber Director, Executive Office of the President, 2024.
[11] E. Ernerfeldt. "egui: an easy-to-use GUI in pure Rust." https://github.com/emilk/egui (accessed 18 April 2026).
[12] W. Crichton, G. Gray, and S. Krishnamurthi, "A grounded conceptual model for ownership types in rust," *Proceedings of the ACM on Programming Languages,* vol. 7, no. OOPSLA2, pp. 1224-1252, 2023.
[13] K. R. C. Siek, "An Undergraduate Computer Systems Curriculum Using Rust," Virginia Tech, 2025.
[14] F. Strömbäck, L. Mannila, and M. Kamkar, "Pilot study of progvis: A visualization tool for object graphs and concurrency via shared memory," in *Proceedings of the 24th Australasian Computing Education Conference*, 2022, pp. 123-132.
[15] F. Strömbäck, L. Mannila, and M. Kamkar, "A weak memory model in Progvis: Verification and improved accuracy of visualizations of concurrent programs to aid student learning," in

Proceedings of the 22nd Koli Calling International Conference on Computing Education Research, 2022, pp. 1-12.
[16] F. Strömbäck, L. Mannila, and M. Kamkar, "Using Model-Checking and Peer-Grading to Provide Automated Feedback to Concurrency Exercises in Progvis," in *Proceedings of the 25th Australasian Computing Education Conference*, 2023, pp. 11-20.
[17] W. Stallings, *Operating Systems: Internals and Design Principles*. Pearson Education Limited, 2018.
[18] A. S. Tanenbaum, "Modern Operating Systems, 4th Edition," ed: Pearson, 2015.
[19] J. C. Adams, E. R. Koning, and C. D. Hazlett, "Visualizing classic synchronization problems: Dining philosophers, producers-consumers, and readers-writers," in *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, 2019, pp. 934-940.
[20] P. Yosifovich, M. E. Russinovich, A. Ionescu, and D. A. Solomon, Windows Internals, Part 1: System architecture, processes, threads, memory management, and more, 7th Edition. Microsoft Press, Pearson, 2017.
[21] A. Allievi, M. E. Russinovich, A. Ionescu, and D. A. Solomon, *Windows Internals, Part 2, 7th Edition*. Microsoft Press, Pearson, 2021.
[22] K. N. Billimoria, Linux Kernel Programming: A comprehensive and practical guide to kernel internals, writing modules, and kernel synchronization. Packt Publishing Ltd, 2024.
[23] J. Sweller, "Cognitive load theory and educational technology," *Educational technology research and development,* vol. 68, no. 1, pp. 1-16, 2020.

https://doi.org/10.31436/iiumej.v27i2.4405

Constant-Time Bitsliced Rijndael-256 on ARM Cortex-M4: On the Limitations of Fixslicing Beyond AES-128

ANDRIANI ADI LESTARI[1], SURYADI MT[2], KALAMULLAH RAMLI[1*], TEDDY SURYA GUNAWAN[3], ESTI RAHMAWATI AGUSTINA[4], SUSILA WINDARTA[5]

[1]*Department of Electrical Engineering, Faculty of Engineering, Universitas Indonesia, Depok 16424, Indonesia*
[2]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok 16424, Indonesia*
[3]*Department of Electrical and Computer Engineering, Kulliyyah of Engineering, International Islamic University Malaysia (IIUM), Kuala Lumpur 50728, Malaysia*
[4]*National Cyber and Crypto Agency, Depok 16516, Indonesia*
[5]*Department of Cybersecurity, Politeknik Siber dan Sandi Negara, Bogor 16120, Indonesia*

**Corresponding author: kalamullah.ramli@ui.ac.id*

(Received: 15 March 2026; Accepted: 5 May 2026; Published online: 10 May 2026)

ABSTRACT: Wider-block ciphers are increasingly needed in high-volume applications, because 128-bit blocks in modes such as Galois/Counter Mode (GCM) limit each invocation to roughly 64 GiB of plaintext per key-nonce pair, forcing complex re-keying strategies. Rijndael-256, the 256-bit-block variant of Rijndael with a 256-bit key, has therefore attracted renewed interest as a natural wider-block companion to Advanced Encryption Standard (AES). At the same time, 32-bit ARM Cortex-M microcontrollers dominate the IoT and embedded landscape, yet, to the best of our knowledge, no constant-time software implementation of Rijndael-256 targeting this platform has been published. This paper addresses that gap. We present a constant-time bitsliced implementation of Rijndael-256 on the ARM Cortex-M4 and provide a systematic structural analysis explaining why fixslicing, the technique that achieves the best-known AES-128 performance on this platform, becomes suboptimal when applied to Rijndael-256. Specifically, the irregular ShiftRows offsets (0, 1, 3, 4) of Rijndael-256 break the uniform register rotation exploited by fixslicing, requiring eight distinct MixColumns compensation variants instead of four. We demonstrate that these compensation variants cost 3.00× as much as executing an explicit, in-place ShiftRows routing using ARM's bitfield instructions. Our macro-inlined assembly variant achieves 6,199 cycles (193.7 cycles/byte) at `-O2`, including packing and unpacking. We provide benchmarks across five compiler optimization levels, constant-time verification over 10^6 samples via DUDECT (maximum t-statistic well below the vulnerability threshold), and per-component cycle breakdowns, showing that the optimal bitslicing strategy is inherently cipher-specific and architecture-dependent.

ABSTRAK: Sifer blok yang lebih lebar semakin diperlukan dalam aplikasi berisipadu tinggi kerana blok 128-bit dalam mod seperti Galois/Counter Mode (GCM) mengehadkan setiap invokasi kepada kira-kira 64 GiB teks biasa bagi setiap pasangan kunci-nonce, sekali gus memerlukan strategi penukaran kunci yang lebih kompleks. Rijndael-256, iaitu varian Rijndael dengan blok 256-bit dan kunci 256-bit, telah menarik semula perhatian sebagai alternatif pelengkap berblok lebih lebar kepada Advanced Encryption Standard (AES). Pada masa yang sama, mikropengawal ARM Cortex-M 32-bit mendominasi landskap Internet of Things (IoT) dan sistem terbenam. Namun, setakat pengetahuan kami, tiada pelaksanaan perisian Rijndael-256 secara bitslicing dan masa-tetap yang menyasarkan platform ini telah diterbitkan. Makalah ini menangani jurang tersebut. Kami membentangkan pelaksanaan

Rijndael-256 secara bitslicing dan masa-tetap pada ARM Cortex-M4 serta menyediakan analisis struktur yang sistematik bagi menjelaskan mengapa fixslicing, iaitu teknik yang mencapai prestasi terbaik yang diketahui bagi AES-128 pada platform ini, menjadi kurang optimum apabila diterapkan pada Rijndael-256. Secara khusus, anjakan ShiftRows Rijndael-256 yang tidak seragam, iaitu (0, 1, 3, 4), mengganggu putaran daftar seragam yang dimanfaatkan oleh fixslicing, lalu memerlukan lapan varian pampasan MixColumns yang berbeza berbanding hanya empat. Kami menunjukkan bahawa varian pampasan ini memerlukan kos 3.00× lebih tinggi berbanding pelaksanaan penghalaan ShiftRows secara eksplisit di tempat menggunakan arahan bitfield ARM. Varian kod himpunan tersisip-makro kami mencapai 6,199 kitaran, bersamaan 193.7 kitaran/bait, pada tahap pengoptimuman `-O2`, termasuk proses pembungkusan dan penyahbungkusan. Kami turut menyediakan penanda aras merentasi lima tahap pengoptimuman pengkompil, pengesahan masa-tetap ke atas 10^6 sampel menggunakan DUDECT dengan statistik-t maksimum yang jauh di bawah ambang kerentanan, serta pecahan kitaran bagi setiap komponen. Dapatan ini menunjukkan bahawa strategi bitslicing yang optimum adalah khusus kepada sifer dan sangat bergantung pada seni bina perkakasan.

KEYWORDS: *ARM Cortex-M4; bitslicing; constant-time implementation; fixslicing; Rijndael-256.*

1. INTRODUCTION

The AES, based on the Rijndael cipher with a fixed 128-bit block size, has served as the foundation of symmetric cryptography since its standardization by NIST in 2001 [1]. However, the 128-bit block size imposes constraints that have become increasingly relevant for high-volume data processing. Modes of operation such as GCM are subject to per-invocation plaintext length constraints: GCM restricts secure plaintext processing to approximately $2^{39} - 256$ bits (roughly 64 GiB) per key-nonce pair [2, 3].

Beyond this limit, the security guarantees of the mode degrade, necessitating re-keying procedures that add complexity to protocol design. Wider-block ciphers address this by enabling birthday-bound modes such as Synthetic IV and Accordion [4]. These constraints have motivated renewed interest in wider-block ciphers. In August 2024, NIST indicated its intent to vet Rijndael with 256-bit blocks and a single 256-bit key for potential standardization, and in December 2024, formally requested public comments on its plan to develop a draft standard (SP 800-197) [5]. The comment period closed on June 25, 2025, and received submissions from both academic researchers and industry stakeholders [6]. This represents the first significant consideration of wider-block Rijndael variants for standardization.

Throughout this paper, we use "Rijndael-256" as shorthand for the specific Rijndael variant with a 256-bit block and a 256-bit key, which corresponds to $N_b = 8$ and $N_k = 8$, which is the configuration under consideration by NIST. Here, N_b denotes the number of 32-bit words in the block (state), and N_k denotes the number of 32-bit words in the key.

1.1. The 32-bit IoT Landscape

A central question for the practical deployment of Rijndael-256 is whether it can be implemented efficiently on resource-constrained processors. The ARM Cortex-M family of 32-bit microcontrollers is the dominant architecture in Internet of Things (IoT) and embedded systems. According to recent industry analyses, ARM Cortex-M cores command approximately 66% of the overall MCU instruction-set architecture share, and the global installed base of IoT-connected devices is forecast to surpass 40 billion units by 2030 [7, 8].

https://doi.org/10.31436/iiumej.v27i2.4405

These devices serve critical functions across industrial automation, smart grid infrastructure, medical monitoring, and automotive systems. Many operate on batteries for extended periods and process sensitive data that requires strong confidentiality guarantees. If Rijndael-256 is standardized, these devices will need constant-time implementations to resist timing side-channel attacks in protocols such as DTLS, secure firmware update, and sensor-to-gateway encryption.

While 64-bit application processors and emerging RISC-V cores with vector extensions are gaining traction in edge computing, the 32-bit Cortex-M class is expected to remain the dominant platform for constrained IoT endpoints for the foreseeable future [7]. The timing of this work is therefore important: the NIST consideration of Rijndael-256 creates a practical need for implementers to understand whether AES-optimized techniques transfer to wider-block Rijndael on microcontrollers. Existing AES/fixslicing studies primarily target AES-128 or AES-256 and do not quantify the structural cost introduced by Rijndael-256's 4×8 state and irregular ShiftRows offsets on scalar 32-bit ARM hardware. Despite this practical need, the published literature contains no constant-time Rijndael-256 implementation for any 32-bit ARM Cortex-M device. This paper addresses that gap.

1.2. Performance Challenges on 32-bit Hardware

Implementing Rijndael-256 on 32-bit hardware presents challenges beyond those encountered with AES-128. The cipher operates on a 4×8 byte state matrix, doubling the state size relative to AES. On the Cortex-M4, which provides 13 general-purpose registers, the 256-bit state consumes eight registers, leaving minimal headroom for intermediate computations.

The ShiftRows transformation uses offsets $(0, 1, 3, 4)$ for $N_b = 8$, which differ from AES-128's uniform $(0, 1, 2, 3)$ offsets. Recent work by Saarinen [6], submitted as a comment to the NIST standardization process, evaluated Rijndael-256 using the RISC-V Vector Cryptography extension (Zvkn). Saarinen reports that the instruction count for 1-kilobyte encryption is $2.66 \times$ higher than for AES-256, primarily because ShiftRows requires additional byte-shuffle operations that cross 128-bit lane boundaries in each round. Our results confirm that this bottleneck persists on scalar 32-bit hardware, but can be addressed effectively using the Cortex-M4's `BFI` (Bit Field Insert) and `UBFX` (Unsigned Bit Field Extract) bit-field instructions.

1.3. Contributions

Our contributions are as follows:

(1) To the best of our knowledge, we present the first constant-time bitsliced implementation of Rijndael-256 on the ARM Cortex-M4 microcontroller, scoped to single-block, encryption-only, software-only execution without a hardware cryptographic accelerator. Our macro-inlined assembly variant achieves 193.7 cycles/byte (6,199 cycles total) at `-O2`, while our function-call assembly variant achieves 203.7 cycles/byte (6,518 cycles), and our portable C variant achieves 311.6 cycles/byte. All assembly cycle counts include bitsliced state packing and unpacking.

(2) We provide a systematic structural analysis demonstrating why fixslicing [9], the state-of-the-art technique for AES-128, becomes suboptimal for Rijndael-256. We show that the irregular ShiftRows offsets $(0, 1, 3, 4)$ double the fixslicing period from $T = 4$ to $T = 8$, breaking the uniform register rotation that makes fixslicing efficient, and requiring eight distinct MixColumns compensation variants that cost on average $3.00 \times$ more than explicit ShiftRows with `BFI`/`UBFX`.

https://doi.org/10.31436/iiumej.v27i2.4405

(3) We provide comprehensive benchmarks across five compiler optimization levels (`-O0` through `-Os`), including a theoretical instruction-level cost analysis. This analysis reveals that ShiftRows achieves a stall factor of $1.02\times$ (near-theoretical), while MixColumns carries a $3.04\times$ stall penalty due to chained data dependencies. We further provide code size measurements, memory footprint comparisons, and empirical constant-time verification over 10^6 samples using Welford's online algorithm, passing the DUDECT statistical framework [10] with a maximum t-statistic well below the vulnerability threshold.

1.4. Paper Organization

The remainder of this paper is organized as follows. Section 2 reviews the Rijndael-256 specification, bitslicing techniques, and related work. Section 3 presents the analysis of why fixslicing becomes inefficient for Rijndael-256. Section 4 describes our implementation. Section 5 presents experimental results. Sections 6 and 7 provide the comparison with AES-128 and a broader discussion, respectively. Section 8 concludes the paper.

2. Background and Related Work

2.1. Rijndael-256 Specification

Rijndael-256 operates on a 4×8 byte state matrix, with 14 rounds, $N_b = 8$, and $N_k = 8$ [11]. Each round applies four transformations: SubBytes (identical S-box to AES), ShiftRows with row offsets $(0, 1, 3, 4)$, MixColumns over $GF(2^8)$ with the same polynomial as AES, and AddRoundKey. The final round omits MixColumns.

The key schedule for $N_k = 8$ applies SubBytes with RotWord and Rcon when $i \equiv 0 \pmod{N_k}$, where i denotes the word index in the expanded key, and additionally applies SubBytes (without rotation) when $i \equiv 4 \pmod{N_k}$. The expanded key comprises 15 256-bit round keys, totaling 480 bytes.

2.2. Bitslicing

Bitslicing, introduced by Biham for DES [12], represents an n-bit cipher state across n registers, where each register holds one bit position from the data. For Rijndael-256, the 256-bit state maps to eight 32-bit ARM registers. All operations are implemented as Boolean circuits, thereby inherently eliminating data-dependent memory accesses and their associated timing leakage.

The S-box is implemented as a Boolean circuit following Boyar and Peralta's depth-16 minimized design [13], with further optimization for ARM Cortex-M by Schwabe and Stoffelen [14].

2.3. Constant-Time Execution Model

In this paper, “constant-time” means that the implementation contains no data-dependent branches or memory accesses, addressing the cache-timing threat model demonstrated by Bernstein [15] and following the formal constant-time definition of Almeida *et al.* [16]. This work addresses timing side-channels only; power, electromagnetic, and fault attacks are outside the scope of the present implementation. All control flow and memory access patterns are independent of the plaintext, key, and intermediate state values. Residual cycle-count variation (as reported in Section 5) arises from microarchitectural effects, such as instruction-cache line alignment and flash prefetch behavior, rather than from data-dependent operations. The T-table

https://doi.org/10.31436/iiumej.v27i2.4405

implementations execute a fixed sequence of instructions (constant instruction count), but perform table lookups indexed by secret-dependent values; the resulting cache access patterns make their execution time data-dependent, i.e., not constant-time in the sense defined above. DUDECT [10] provides empirical evidence of no detectable timing leakage under the tested conditions; it is not a formal proof of constant-time execution.

2.4. Fixslicing for AES-128

Fixslicing, introduced by Adomnicai *et al.* [9, 17], absorbs the ShiftRows permutation into MixColumns by pre-rotating round keys to compensate for the accumulated ShiftRows offset. For AES-128 ($N_b = 4$, offsets $(0, 1, 2, 3)$), ShiftRows has period $T = 4$, meaning four distinct MixColumns variants suffice. On the Cortex-M4, each variant reduces to word-level rotations handled at zero cost by the barrel shifter. This approach achieves approximately 80 cycles/byte for two-block parallel AES-128 encryption [9].

2.5. ARM Cortex-M4 Architecture

The ARM Cortex-M4 is a 32-bit processor based on the ARMv7E-M architecture, featuring a single-cycle barrel shifter [18]. The `BFI` and `UBFX` instructions enable byte extraction and insertion in 1–2 cycles [18]. The STM32F407 variant used in our experiments features 1 MB of flash, 192 KB of SRAM, and an ART accelerator for zero-wait-state flash access at 168 MHz [19].

2.6. Related Work

Constant-time AES implementations on Cortex-M4 have been studied extensively. Schwabe and Stoffelen [14] presented an optimized bitsliced implementation achieving approximately 100 cycles/byte. Adomnicai and Peyrin improved this to approximately 80 cycles/byte using fixslicing [9], processing two 128-bit blocks in parallel.

For Rijndael variants beyond AES-128, the published literature is sparse. Saarinen's evaluation of RISC-V Vector architectures [6] represents the most relevant contemporary work, but targets a different hardware class. Constant-time software implementations for embedded 32-bit processors have not been previously reported.

3. WHY FIXSLICING BECOMES INEFFICIENT FOR RIJNDAEL-256

3.1. The Period Theorem

Fixslicing defers ShiftRows and compensates within MixColumns. After k rounds without explicit ShiftRows, the accumulated row offsets are $(0, k \cdot s_1, k \cdot s_2, k \cdot s_3) \bmod N_b$, where s_1, s_2, s_3 denote the non-zero ShiftRows offsets. The technique requires these offsets to return to $(0, 0, 0, 0)$, establishing a period T.

Theorem 1 (ShiftRows Period). For Rijndael with block size N_b and non-zero ShiftRows offsets s_1, s_2, s_3, the ShiftRows period is

$$T = \operatorname{lcm}\left(\frac{N_b}{\gcd(N_b, s_1)}, \frac{N_b}{\gcd(N_b, s_2)}, \frac{N_b}{\gcd(N_b, s_3)}\right). \quad (1)$$

Proof. Row i, with offset s_i, returns to its original position after k rounds when $k \cdot s_i \equiv 0 \pmod{N_b}$. Thus k must be a multiple of $N_b/\gcd(N_b, s_i)$ for each non-zero row. The period T is the smallest k satisfying all three congruences simultaneously, giving the least common multiple above. ■

https://doi.org/10.31436/iiumej.v27i2.4405

Definition 1 (Bitsliced State Packing). Let $W = 32$ denote the register width in bits and N_b the number of state columns, and $p = 8/N_b$ the number of parallel blocks in the layout. A bitsliced Rijndael state packs each bit-plane into a W-bit register. The register is partitioned into four $(W/4)$-bit row groups, and each column occupies a field of width $g = 4 \cdot p = W/N_b$ bits, distributed across the four row groups and not contiguous in the full register. For AES-128, $p = 2$ and $g = 8$ bits (byte-aligned). For Rijndael-256, $p = 1$ and $g = 4$ bits (sub-byte).

For AES-128 ($N_b = 4$, offsets $1, 2, 3$), $T = \mathrm{lcm}(4, 2, 4) = 4$. For Rijndael-256 ($N_b = 8$, offsets $1, 3, 4$), $T = \mathrm{lcm}(8, 8, 2) = 8$. To illustrate the practical difference, after two rounds, the accumulated offsets for AES-128 are $(0, 2, 0, 2)$, whereas those for Rijndael-256 are $(0, 2, 6, 0)$. Because each column in the two-block AES-128 layout spans $g = 8$ bits, the shifts are byte-aligned and can be performed by the ARM barrel shifter at zero cost. For Rijndael-256, each column spans only $g = 4$ bits, so column shifts are sub-byte and require per-row bit-field manipulation via `UBFX/BFI` rather than barrel-shifted rotation.

Lemma 1 (Rotation Granularity Constraint). On the ARMv7-M architecture, the barrel shifter applies a uniform rotation to an entire 32-bit word. The accumulated ShiftRows compensation for row i at phase k requires a rotation by $r_i(k) = (k \cdot s_i \bmod N_b) \cdot p$ bits within that row's $(W/4)$-bit row group. This per-row rotation can be absorbed into a single barrel-shifted data-processing instruction if and only if all the required rotations are compatible with a whole-word, byte-aligned rotation; otherwise, the rows require independent sub-word manipulation.

$$\text{Condition (2): } r_1(k) = r_2(k) = r_3(k), \text{ with byte-aligned } r_i(k). \tag{2}$$

When Condition (2) does not hold, each non-aligned row requires an extract–shift–insert sequence (`UBFX, LSL/LSR, BFI`), costing at least 3 instructions per row.

Proof. The barrel shifter in `EOR Rd, Rn, Rm, ROR #n` applies a uniform rotation to the entire 32-bit operand Rm. For this to implement the correct compensation, the rotation amount must match the required row-wise rotations without splitting the $(W/4)$-bit row groups. When the amounts differ, or when the rotation is sub-byte as in Rijndael-256, the rows occupy disjoint $(W/4)$-bit row-groups within the register, and no single word-level operation can independently rotate sub-word fields on ARMv7-M. Each misaligned row, therefore, requires at minimum one `UBFX` to extract the $(W/4)$-bit row-group, one shift to rotate it, and one `BFI` to insert it back. Conversely, when Condition (2) holds and $r_i(k)$ is byte-aligned, the uniform rotation by $r_i(k)$ bits correctly compensates all rows simultaneously via a single barrel-shifted instruction, as the rotation amount is a multiple of $(W/4)$ and preserves row-group boundaries. ■

Theorem 2 (Phase-wise Compensation Cost). Let $a_i(k) = k \cdot s_i \bmod N_b$ denote the accumulated column offset for row i at phase k. Define the number of rows requiring compensation at the phase k as

$$\mu(k) = |\{i \in \{1,2,3\} : a_i(k) \neq 0\}|. \tag{3}$$

By Lemma 1, in the row-grouped layout, each non-zero row resides in a distinct row-group within the register and therefore requires its own extract-shift-insert sequence; rows that happen to share the same offset value cannot share a single `UBFX/BFI` pair because the row-groups are not contiguous. Each such sequence costs at least $C_{ins} = 3$ instructions. For Rijndael-256 with offsets (1,3,4), we have $\mu(k) \geq 1$ for all $k \in \{1, \ldots, 7\}$, since $k < T$

https://doi.org/10.31436/iiumej.v27i2.4405

guarantees at least one $a_i(k) \neq 0$. Over one full period $T = 8$, the total additional instruction count is at least

$$C_{ins} \cdot \sum_{k=0}^{T-1} \mu(k) \geq C_{ins}(T-1). \quad (4)$$

Proof. For any $k \in \{1, \ldots, T-1\}$, since $k < T$ and $s_i \neq 0$, we have $k \cdot s_i \not\equiv 0 \pmod{N_b}$ for at least one $i \in \{1, 2, 3\}$, hence $\mu(k) \geq 1$. Summing over all phases gives $\sum_{k=0}^{T-1} \mu(k) \geq T - 1$, and therefore the total additional instruction count is at least $C_{ins}(T-1)$. For Rijndael-256 specifically, we enumerate the offsets $(a_1(k), a_2(k), a_3(k)) \bmod 8$ for $k = 0, \ldots, 7$ with $(s_1, s_2, s_3) = (1, 3, 4)$ and $N_b = 8$:

k	$(\boldsymbol{a_1}, \boldsymbol{a_2}, \boldsymbol{a_3})$	$\mu(k)$	Min. Extra instr.
0	(0, 0, 0)	0	0
1	(1, 3, 4)	3	9
2	(2, 6, 0)	2	6
3	(3, 1, 4)	3	9
4	(4, 4, 0)	2	6
5	(5, 7, 4)	3	9
6	(6, 2, 0)	2	6
7	(7, 5, 4)	3	9

Summing the last column gives 54 instructions, which is the tight enumeration-based lower bound for this row-grouped layout. The coarser bound $C_{ins}(T-1) = 21$ instructions are only a conservative universal bound and is not used for the Rijndael-256 cost claim. For even k, the third offset is zero because $4k \equiv 0 \ (mod\ 8)$, but the first two offsets remain non-zero, so $\mu(k) = 2$; for odd k, all three offsets are non-zero, so $\mu(k) = 3$. ■

Corollary 1 (Asymptotic Lower Bound). For any Rijndael variant with period $T > 1$, the fixsliced MixColumns overhead is at least $C_{ins} \cdot (T-1) \cdot \lfloor r/T \rfloor = \Omega(r)$ instructions over the full cipher, where C_{ins} denotes the minimum instruction cost of one extract-shift-insert compensation and r denotes the number of rounds. This overhead is absent in the canonical bitsliced implementation, which realizes ShiftRows through explicit `BFI/UBFX` byte routing rather than phase-dependent MixColumns compensation.

Proof. By Theorem 2, phase k contributes at least $C_{ins} \cdot \mu(k)$ instructions, so a full period contributes at least $C_{ins} \cdot \sum_{k=0}^{T-1} \mu(k)$. Since every non-identity phase admits at least one non-zero offset, we have $\mu(k) \geq 1$ for $k \in \{1, \ldots, T-1\}$, yielding the conservative per-period lower bound $C_{ins} \cdot (T-1)$. The r rounds of the cipher decompose into $\lfloor r/T \rfloor$ disjoint periods of length T, together with at most $T-1$ residual rounds whose contribution we discard. Summing the per-period bound over these $\lfloor r/T \rfloor$ disjoint periods, therefore gives a full-cipher lower bound of $C_{ins} \cdot (T-1) \cdot \lfloor r/T \rfloor$, and since T is fixed this quantity is $\Omega(r)$. For Rijndael-256, the explicit enumeration gives $\sum_{k=0}^{T-1} \mu(k) = 18$, hence 54 instructions per period and $54 \cdot \lfloor r/T \rfloor$ instructions over the full cipher. In the canonical bitsliced approach, ShiftRows is implemented by explicit `BFI/UBFX` routing in a single semi-fused routine, so no phase-dependent MixColumns compensation is emitted. ■

Proposition 1 (Semi-fused vs. Fixsliced Comparison). The semi-fused canonical approach, which performs ShiftRows and MixColumns explicitly in a single routine, costs 297 cycles per round at `-O2` on the STM32F407. The fixsliced approach eliminates ShiftRows but replaces the canonical MixColumns with phase-dependent compensated variants averaging 891 cycles per round (Table 1), a ratio of 3.00 ×. The net balance is negative.

https://doi.org/10.31436/iiumej.v27i2.4405

Proof. On the STM32F407 at `-O2`, we measure each approach as a single per-round unit (Table 1): the semi-fused canonical routine costs C_{SF} = 297 cycles, while the fixsliced compensated MixColumns averages $\bar{C}_{FS}$ = 891 cycles across all eight phases. Since $\bar{C}_{FS}/C_{SF} = 891/297 = 3.00 > 1$, the fixsliced approach is strictly more expensive per round. Over the full 14-round cipher, the measured total encryption cost corroborates the 3.00 × overhead ratio: 11,633 cycles (fixsliced hardcoded) versus 9,971 cycles (semi-fused), a difference of 1,662 cycles or 16.7%. ■

Table 1. Per-phase fixsliced MixColumns cost for Rijndael-256 on STM32F407 (cycles, `-O2`).

Phase*	Cycles	vs Semi-fused
Semi-fused	297	1.00× (base)
Phase 0	949	3.19×
Phase 1	805	2.71×
Phase 2	949	3.19×
Phase 3	853	2.87×
Phase 4	949	3.19×
Phase 5	853	2.87×
Phase 6	949	3.19×
Phase 7	821	2.76×
Average	891	3.00×

*Phase i in this table, corresponds to the mathematical phase $k = (i + 1) \bmod 8$ from Theorem 2. The identity phase ($k = 0$, zero accumulated offset) appears as Phase 7.

Note on the 3.00 × *ratio.* Both implementations are compiled in C at `-O2`; the fixsliced variant decomposes each compensated MixColumns into three explicit steps, undoing the accumulated offset (Step A), performing canonical MixColumns (Step B), applying the next-phase offset (Step C), rather than fusing the compensation into the MixColumns XOR tree. For Rijndael-256, such fusion would not reduce the instruction count: the sub-byte compensation ($g = 4$) cannot be absorbed by the barrel shifter, so each non-zero row still requires an `UBFX/BFI` sequence regardless of whether it is performed inline or as a separate pass. The 3.00 × ratio therefore reflects both the algorithmic overhead (two rotation passes versus one) and compiler overhead (generic C rotation versus hand-scheduled assembly). At the instruction level, the hand-tuned assembly ShiftRows costs 96 instructions per call (Table 8), a fixsliced round would require two such passes plus MixColumns, giving a theoretical lower bound of $96 + 96 + 27 = 219$ instructions versus $96 + 27 = 123$ for the canonical round, a ratio of 1.78 ×, confirming that the overhead is structural, not an artifact of the C implementation.

Remark 1 (Decryption Symmetry). The preceding results apply equally to decryption. InvShiftRows uses offsets $(0, N_b - s_1, N_b - s_2, N_b - s_3)$; for Rijndael-256 these yield $(0, 7, 5, 4)$. The inverse period is $T_{inv} = \mathrm{lcm}\left(\frac{8}{\gcd(8,7)}, \frac{8}{\gcd(8,5)}, \frac{8}{\gcd(8,4)}\right) = \mathrm{lcm}(8, 8, 2) = 8 = T$. The column field width g is unchanged, and the same byte-alignment condition determines efficiency. Therefore, separate proofs for InvShiftRows are unnecessary.

3.2. General Condition for Fixslicing Inefficiency

The preceding analysis of Rijndael-256 is a special case of a broader structural limitation. We now characterize, for the general Rijndael family, the conditions under which fixslicing becomes more expensive than the canonical approach. The primary bottleneck is not the

https://doi.org/10.31436/iiumej.v27i2.4405

specific ShiftRows offset values: all Rijndael variants with $N_b \in \{4, 5, 6\}$ share the same ShiftRows offsets $(1, 2, 3)$; $N_b = 7$ uses $(1, 2, 4)$; only $N_b = 8$ uses $(1, 3, 4)$. Rather, the decisive factor is the column field width g. When $g < 8$, column shifts correspond to sub-byte rotations within each $(W/4)$-bit row-group, which no single ARMv7-M data-processing instruction can absorb for free.

Theorem 3 (Fixslicing Efficiency Threshold). Let C_{SR} be the cost (in instructions) of an explicit ShiftRows per round, and let C_{ins} be the cost of a single extract-shift-insert operation. Fixslicing incurs zero ShiftRows compensation cost within MixColumns if

(C1) Byte-alignment: $g = W/N_b \geq 8$, i.e., each column occupies at least one full byte in the 32-bit register layout, so a column shift corresponds to a byte-aligned rotation that can be absorbed by the barrel shifter's second operand at zero additional instruction cost.

In particular, AES-128 satisfies (C1) because $N_b = 4$ and $g = 8$. Its fixsliced MixColumns variants absorb all compensation rotations into barrel-shifted operands of existing `EOR` instructions, despite the four rows sharing each register.

When (C1) is violated (g < 8), the per-row compensation of $(k \cdot s_i \bmod N_b) \cdot p$ bits is sub-byte and requires extract–shift–insert sequences. A sufficient (but not necessary) condition for fixslicing to be more expensive than the canonical approach is:

$$\frac{C_{ins}}{T} \sum_{k=0}^{T-1} \mu(k) > C_{SR}\,, \tag{5}$$

where $\mu(k)$ is the number of rows requiring compensation at phase k (Theorem 2). The left-hand side is a lower bound on the average per-round compensation cost; the actual overhead may be substantially larger due to factors not captured by the instruction count (compiler-generated bit-manipulation code expanding beyond the minimum 3-instruction pattern, pipeline stalls from chained data dependencies, function-dispatch, and instruction-cache pressure from the T-variant phase table). Hence, failure to satisfy (5) does not imply that fixslicing wins. For borderline variants, the verdict requires empirical measurement.

Proof. Sufficiency of (C1). When $g \geq 8$, a column shift corresponds to a register-level rotation. Adomnicai and Peyrin [9] exploit this property by pre-permuting round keys offline (during key expansion) so that, after the accumulated ShiftRows shift, the bits required at each register position are already laid out correctly in the round key. At runtime, the rotation is applied to the pre-permuted round key as the second operand of an `EOR Rd, Rn, Rm, ROR #sh` instruction, not to the state itself, where $sh = g \cdot a_i(k) = g \cdot (k \cdot s_i \bmod N_b)$ denotes the rotation amount in bits for each row $i \in \{1, 2, 3\}$. The ARMv7-M barrel shifter absorbs this rotation at zero instruction cost. Each fixsliced phase k emits zero additional compensation instructions.

Cost lower bound when (C1) fails. If $g < 8$, the corresponding rotation amount is sub-byte and can split bits across row-group boundaries, corrupting the per-row data. Since ARMv7-M lacks sub-word rotation instructions that operate independently on g-bit fields, each non-zero row compensation requires an extract-shift-insert sequence `(UBFX, shift, BFI)`, costing at least $C_{ins} = 3$ instructions.

Sufficient condition for Canonical to win. When compensation is required, by Lemma 1 each non-zero row contributes at least C_{ins} instructions, so phase k adds at least $C_{ins} \cdot \mu(k)$ instructions to MixColumns. Averaged over one period T, the per-round overhead is therefore at least $(C_{ins}/T) \cdot \sum_k \mu(k)$. Fixslicing saves C_{SR} per round by eliminating explicit ShiftRows. Therefore, if $C_{SR} < (C_{ins}/T) \cdot \sum_k \mu(k)$, the canonical approach is provably cheaper. The

converse does not hold, because real implementations may incur overhead far above the lower bound. ■

Remark 2 (Rijndael-256 Lies in the Borderline Regime). For Rijndael-256, $\sum_k \mu(k) = 18$ (Theorem 2), giving a formal lower bound of $3 \cdot 18/8 = 6.75$ instructions per round. The C ShiftRows implementation costs approximately 219 cycles per round at `-O2` (Table 7), while the hand-tuned assembly routing costs 96 instructions and measures 98–168 cycles depending on optimization context (Tables 7 and 8). Because these values substantially exceed the formal lower bound, Theorem 3 must be read as a conservative sufficient condition only. It does not predict Rijndael-256's actual outcome. The operational verdict for Rijndael-256 is delivered by Proposition 1, which shows by direct measurement that the canonical semi-fused approach is faster by 1,662 cycles in the C-only comparison.

Remark 3 (Rijndael Variant Analysis). Table 2 applies this criterion to all Rijndael block lengths. All variants with $N_b \in \{4, 5, 6\}$ share the same ShiftRows offsets (1,2,3); $N_b = 7$ uses $(1, 2, 4)$; and $N_b = 8$ uses $(1, 3, 4)$. Despite identical offsets, fixslicing is efficient only for AES-128 ($N_b = 4$), because it is the only variant where $g \geq 8$. For $N_b \in \{5, 6, 7\}$, the column field width g is not even an integer, meaning that state columns cannot be packed into byte-aligned fields within 32-bit registers. This represents a more severe obstacle than the sub-byte fields of Rijndael-256 ($g = 4$).

Table 2. Fixslicing feasibility for all Rijndael variants ($W = 32$, p parallel blocks as defined).

Variant	N_b	(s_1, s_2, s_3)	T	p	g	Verdict
Rijndael-128 (AES)	4	(1,2,3)	4	2.0	8.0	Efficient
Rijndael-160	5	(1,2,3)	5	1.6†	6.4†	Not applicable
Rijndael-192	6	(1,2,3)	6	1.3†	5.3†	Not applicable
Rijndael-224	7	(1,2,4)	7	1.1†	4.6†	Not applicable
Rijndael-256	8	(1,3,4)	8	1.0	4.0	Inefficient

† Variants with non-integer g and p cannot be evenly packed into 32-bit row-groups, so the standard bitsliced layout is not realizable for these variants.

3.3. Measured Per-Phase Costs

Figure 1 illustrates why fixslicing succeeds for AES-128 but fails for Rijndael-256. Table 1 uses dispatch labels in which the Phase i maps to the mathematical phase $k = (i + 1) \bmod 8$. Therefore, phases requiring compensation for all three rows, $k \in \{1, 3, 5, 7\}$, correspond to Phases 0, 2, 4, and 6 in Table 1 and are the most expensive at 949 cycles each, consistent with Theorem 2's prediction. Phase 7 corresponds to $k = 0$, the identity case, and costs 821 cycles, essentially representing canonical MixColumns plus dispatch overhead with no compensation. Phases with $\mu(k) = 2$ range from 805 to 853 cycles depending on the non-zero shift amounts. Even the cheapest compensated phase costs 2.71 × the semi-fused baseline.

https://doi.org/10.31436/iiumej.v27i2.4405

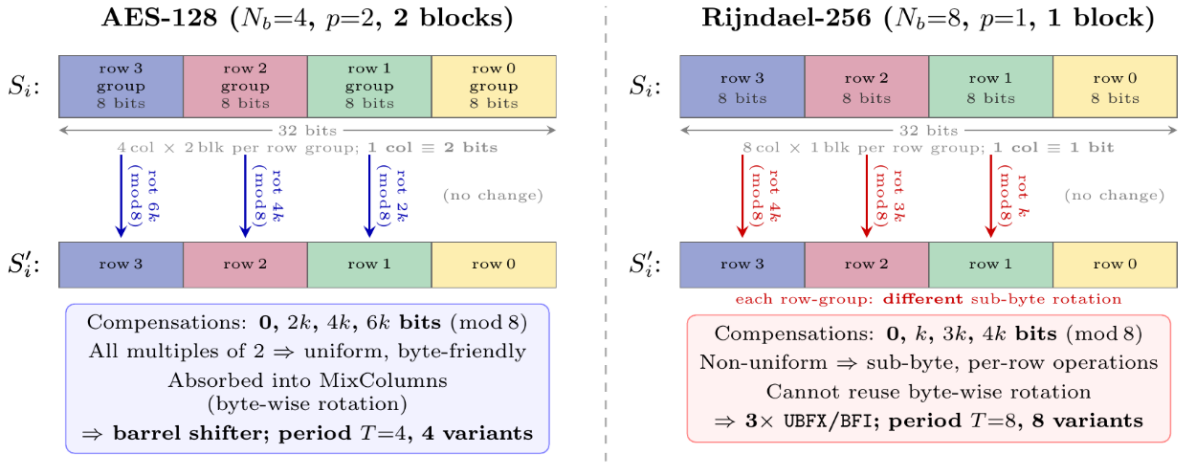


Figure 1. Fixslicing ShiftRows compensation for AES-128 ($g = 8$, byte-aligned) vs. Rijndael-256 ($g = 4$, sub-byte `UBFX/BFI`).

3.4. Key Schedule Overhead

The fixsliced key schedule must pre-rotate round keys to match each phase's accumulated offset. Table 3 shows that this costs 6,370 cycles versus 3,285 cycles for the canonical bitsliced schedule at `-O2`, a ratio of 1.94 × . At `-Os`, the ratio increases to 3.06 × because the rotation code sequences resist size optimization, causing instruction cache pressure.

Table 3. Key schedule cost for Rijndael-256 on STM32F407 (cycles).

Key Schedule	-O0	-O1	-O2	-O3	-Os
BS canonical	21,231	5,085	3,285	3,267	3,234
FS fixsliced	40,570	9,829	6,370	6,189	9,894
FS/BS ratio	1.91	1.93	1.94	1.89	3.06

4. IMPLEMENTATION

We describe the key components of our implementation. Table 4 summarizes the notation used for the implementation variants discussed throughout this paper.

Table 4. Implementation variants and abbreviations.

Abbreviation	Description
Full ASM (macro)	Hand-tuned assembly, macro-inlined (S-box + MC + SR)
Full ASM (func)	Hand-tuned assembly, function-call (S-box + MC + SR)
Inline ASM SR	C bitsliced with inline ASM ShiftRows
BS C (semi-f.)	Pure C bitsliced, explicit ShiftRows
FS hardcoded	Fixsliced with 8 hardcoded MC variants
FS generic	Fixsliced with generic MC compensation
T-table 4T	Four lookup tables (4×1 KB)
T-table 1T	One lookup table (1 KB) + barrel shifter

4.1. Bitsliced State Layout

We adopt a byte-based packing in which each 32-bit register byte holds one state row across eight columns, with reversed column order (column 0 in the MSB, column 7 in the LSB). In this representation, `ROR32(x,8)` shifts data between rows, and ShiftRows requires left rotation per byte within each register. Figure 2 illustrates this mapping.

https://doi.org/10.31436/iiumej.v27i2.4405

	row 3 [31:24]					row 0 [7:0]			
	c_0	c_1	$\cdots$	c_7	$\cdots$	c_0	c_1	$\cdots$	c_7
S_0	$b_{3,0}[0]$	$b_{3,1}[0]$	$\cdots$	$b_{3,7}[0]$	$\cdots$	$b_{0,0}[0]$	$b_{0,1}[0]$	$\cdots$	$b_{0,7}[0]$
S_1	$b_{3,0}[1]$	$b_{3,1}[1]$	$\cdots$	$b_{3,7}[1]$	$\cdots$	$b_{0,0}[1]$	$b_{0,1}[1]$	$\cdots$	$b_{0,7}[1]$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$		$\vdots$
S_7	$b_{3,0}[7]$	$b_{3,1}[7]$	$\cdots$	$b_{3,7}[7]$	$\cdots$	$b_{0,0}[7]$	$b_{0,1}[7]$	$\cdots$	$b_{0,7}[7]$

Figure 2. Rijndael-256 bitsliced register layout. Register S_i holds bit i of each state byte. Cell $b_{r,c}[i]$ equals plaintext bit $(4c + r) \cdot 8 + i$.

Figure 3 shows the exact bit-level mapping. The 3-stage `swpmv` butterfly establishes a reversed column order: column 0 maps to bit 7 (MSB) of each row group, and column 7 to bit 0 (LSB), so that a left rotation `(ROL)` within an 8-bit group directly implements ShiftRows. The precise packing rule is $S_i[8r + (7 - c)] = b_{r,c}[i]$.

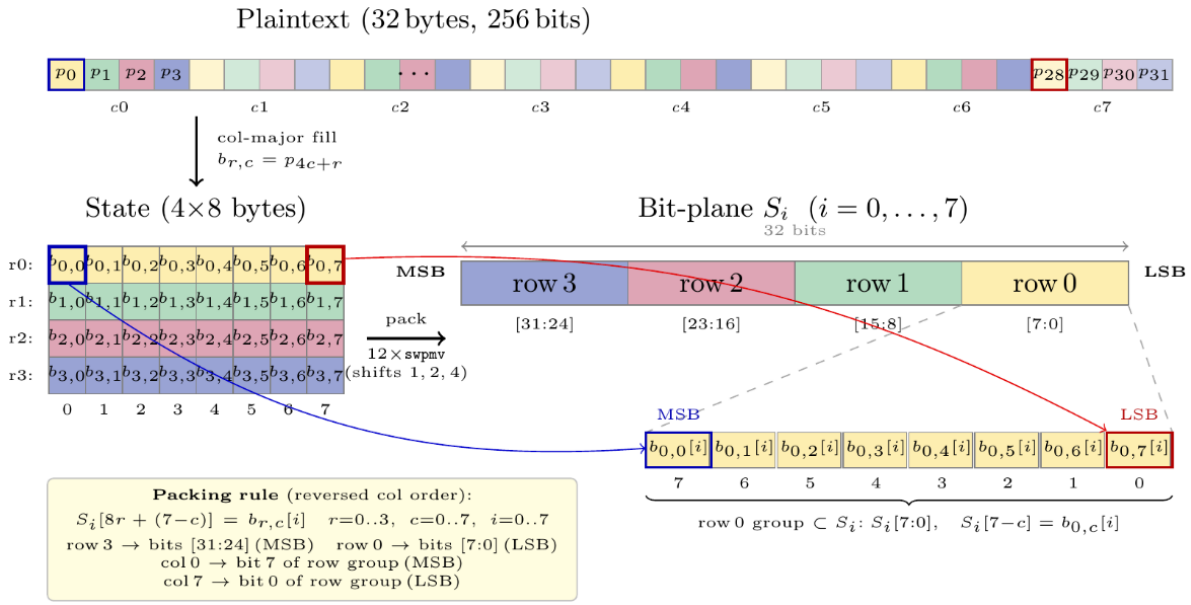


Figure 3. Bit-level packing layout. The 3-stage `swpmv` butterfly maps each state byte $b_{r,c}$ into bit-plane S_i at position $8r + (7 - c)$. Row 0 occupies the LSB byte [7:0]; row 3 the MSB byte [31:24].

4.2. SubBytes

The S-box uses the Schwabe–Stoffelen Boolean circuit [14], which operates on eight bit-plane registers via Boolean operations without memory accesses. Care must be taken not to duplicate NOT operations in the key schedule, as the Boyar–Peralta circuit [13] already incorporates them internally.

4.3. ShiftRows

We implement ShiftRows in two ways: pure C with byte rotation macros, and inline ARM assembly using `BFI/UBFX` instructions. The inline assembly variant extracts bytes using `UBFX`, rotates them according to the offsets $(0, 1, 3, 4)$, and reinserts them using `BFI`. At `-O3`,

https://doi.org/10.31436/iiumej.v27i2.4405

the assembly variant costs 98 cycles versus 260 for C (2.65 × improvement). This approach addresses the byte-shuffle bottleneck identified by Saarinen [6] in the RISC-V context, using the Cortex-M4's bit-field instructions rather than vector permutations.

4.4. MixColumns

Multiplication by 2 in $GF(2^8)$ at MixColumns is a left shift with conditional XOR of `0x1B` (denoted `xtime`), and multiplication by 3 is `xtime` XOR with the original. In our bitsliced design, the MixColumns step follows the implementation approach originally proposed by Käsper and Schwabe [20] for x86, adapted for ARM by Schwabe and Stoffelen [14], and extended to the eight-column Rijndael-256 state. Multiplication by 2 in $GF(2^8)$ is realized by rotating the eight bit-plane registers and XORing the high-bit plane into the appropriate reduction planes; multiplication by 3 adds the original plane to the doubled result. All operations are unconditional register-level `EOR` and `ROR` instructions with no data-dependent branches or memory accesses, preserving the constant-time property.

For the fixsliced variants (FS hardcoded, FS generic), the same clean MixColumns circuit is used internally, but wrapped with per-byte compensation rotations. Each phase-k variant applies three steps: (A) rotate each byte of the eight-bit plane registers to undo the accumulated ShiftRows offset, (B) execute the clean MixColumns, and (C) rotate bytes again to impose the next phase's offset. The rotation amounts are pre-computed per phase in the hardcoded variant; the generic variant computes them at runtime. Two phases incur reduced compensation cost: phase 1 requires no Step A, and phase 7 requires no Step C, as their respective rotation amounts are all zero. The additional cost relative to the canonical circuit arises entirely from Steps A and C, which require approximately $3 \cdot \mu(k)$ instructions per step, giving $6 \cdot \mu(k)$ total across Steps A and C for $\mu(k)$ non-zero rows, consistent with the overhead analyzed in Section 3.

4.5. Full Assembly Variants

Our highest-performance variants combine the Schwabe–Stoffelen S-box [14], the canonical bitsliced MixColumns described in Section 4.4, and custom `BFI/UBFX` ShiftRows for the $(0, 1, 3, 4)$ offsets. We provide two versions that expose a trade-off between code size and performance.

The *function-call variant* (Full ASM (func)) wraps the round function in a standard ARM procedure. It achieves 6,518 cycles (203.7 cycles/byte) at `-O2`, is compiler-independent, and requires 658 bytes of flash for the ASM cipher core.

The *macro-inlined variant* (Full ASM (macro)) expands the round function at each of the 14 call sites, eliminating all `BL/BX` branching and register save/restore overhead. This reduces the per-round overhead by approximately 23 cycles, yielding 6,199 cycles (193.7 cycles/byte) at the cost of a larger flash footprint. Both variants produce cycle counts that are essentially invariant across compiler optimization flags, since all inner-loop code is hand-written assembly.

4.6. T-table Baseline

For comparison, we implement 4-table and 1-table T-table variants (not in constant time). The standard relationship $T_j[x] = ROL32(T_0[x], 8j)$ requires left rotation in practice. Concretely, $T_1 = ROL32(T_0, 8)$, $T_2 = ROL32(T_0, 16)$, and $T_3 = ROL32(T_0, 24)$. On the Cortex-M4, the barrel shifter absorbs this at zero cost. On RISC-V RV32I, which lacks a native rotation instruction, this would add 2–3 cycles per derivation, potentially affecting the relative advantage of 1-table versus 4-table strategies.

https://doi.org/10.31436/iiumej.v27i2.4405

5. EXPERIMENTAL RESULTS

All benchmarks were conducted on the STM32F407VGT6 Discovery Board, with the ART Accelerator enabled and the core clock set to 168 MHz. Cycle counts were measured using the DWT CYCCNT register with interrupts disabled. The evaluation firmware was compiled with the GNU Arm Embedded GCC toolchain (arm-none-eabi-gcc; exact release recorded in the build artifact) using the optimization flags reported in the tables (`-O0`, `-O1`, `-O2`, `-O3`, and `-Os`) and targeting ARMv7E-M/Thumb-2. For the assembly variants, all reported cycle counts include bitsliced state packing and unpacking, as these operations are integrated into `rj256_encrypt_asm`. Source code, build scripts, and raw measurement logs are available from the corresponding author upon reasonable request and will be archived publicly upon acceptance.

5.1. Encryption Performance

Table 5 presents encryption cycle counts across all optimization levels. The macro-inlined assembly variant achieves 6,199 cycles (193.7 cycles/byte), while the function-call variant achieves 6,518 cycles (203.7 cycles/byte). Both are essentially invariant across compiler optimization flags. The semi-fused C variant achieves 9,971 cycles (311.6 cycles/byte) at `-O2`, outperforming the fixsliced hardcoded variant (11,633 cycles, 363.5 cycles/byte) by 16.7%.

Among non-constant-time implementations, the 1-table variant achieves 3,252 cycles (101.6 cycles/byte) at `-O2`, outperforming the 4-table variant by 20%. The barrel shifter makes `ROR` at runtime effectively free, while the single 1 KB table is more cache-friendly than four 1 KB tables. The `-O2` level produces the best results for both bitsliced implementations and the T-table 1T variant; for the T-table 4T variant, `-O1` is optimal. At `-O3`, the bitsliced C variant regresses by 6% due to aggressive inlining causing code bloat.

Table 5. Encryption performance for Rijndael-256 on STM32F407 at 168 MHz (cycles).

Implementation	-O0	-O1	-O2	-O3	-Os
Constant-time:					
Full ASM (macro)	6,199	6,199	6,199	6,199	6,199
Full ASM (func)	6,524	6,521	6,518	6,519	6,517
Inline ASM SR	42,064	9,044	9,102	8,295	9,635
BS C (semi-f.)	48,028	10,129	9,971	10,559	11,829
FS hardcoded	66,972	11,719	11,633	12,322	20,527
FS generic	76,004	19,308	17,736	15,084	23,384
Non-constant-time:					
T-table 4T	8,624	3,800	4,036	4,300	3,968
T-table 1T	8,925	3,891	3,252	3,842	3,801

5.2. Timing Variability Analysis

Table 6 reports the timing range (max-min) over 10^6 iterations. The constant-time implementations range from 46–140 cycles, while T-table variants range from 132–180 cycles. The constant-time implementations contain no data-dependent memory accesses; the residual variability arises solely from instruction-fetch effects (cache line alignment, flash prefetch). The T-table variability is systematically higher because different plaintext values access different cache lines, constituting exploitable timing leakage.

https://doi.org/10.31436/iiumej.v27i2.4405

Table 6. Timing variability for Rijndael-256 on STM32F407 (cycle range over 10^6 iterations).

Implementation	-O0	-O1	-O2	-O3	-Os
Full ASM (func)	79	62	49	46	48
BS C (semi-f.)	140	132	70	57	118
FS hardcoded	101	86	93	68	121
FS generic	91	68	61	60	76
T-table 4T	172	148	151	174	132
T-table 1T	180	145	150	172	135

5.3. Component Breakdown

Table 7 presents the per-component cycle costs across all optimization levels. SubBytes dominates at 293 cycles per call at `-O2`, consistent with the complexity of the Boolean S-box circuit. ShiftRows is the second most expensive component. The inline assembly variant provides a 23% improvement over C (168 vs. 219 cycles). MixColumns is efficient at 82 cycles. Packing and unpacking together consume 332 cycles, a one-time overhead amortized in modes processing multiple blocks.

Table 7. Per-component cycle costs for the bitsliced Rijndael-256 implementation on STM32F407 (cycles per call).

Component	-O0	-O1	-O2	-O3	-Os
Packing	990	209	109	109	108
SubBytes	1,500	297	293	293	304
ShiftRows (C)	997	220	219	260	332
ShiftRows (ASM)	381	156	168	98	155
MixColumns	357	83	82	82	85
Unpacking	1,115	217	223	222	219

5.4. Theoretical Instruction-Level Cost Analysis

We analyze the theoretical cycle cost of the macro-inlined assembly variant by categorizing instruction sequences by cryptographic role and computing the minimum cycle counts under the assumption of zero pipeline stalls. Comparing these theoretical minima against measured results quantifies the pipeline stall overhead per component, revealing which operations are the primary performance bottlenecks on the Cortex-M4.

5.4.1. Cycle Model

We apply the following Cortex-M4 single-issue cycle model: data-processing instructions (`EOR, AND, ORR, MOV, ROR,` with or without an inline barrel shift) execute in 1 cycle; `UBFX` and `BFI` execute in 1 cycle; `LDR/STR` cost 2 cycles; `LDMIA/STMIA` for n registers cost $n + 1$ cycles; `PUSH/POP` for n registers cost $n + 1$ cycles. Load-use hazards add 1 stall cycle each.

5.4.2. Per-Component Analysis

Packing: The 12 `swpmv` macro invocations (each expanding to 4 data-processing instructions) plus the function prologue (`PUSH` {14 regs}, `SUB`, `STR`, `MOV`, `LDM` {8 regs}) and mask-setup (`MOVW`, `MOVT`, 2×`EOR`) contribute 80 theoretical cycles.

AddRoundKey (ARK): Each round loads two groups of four round-key words (`LDR` + 2×`LDMIA` {4}) and (8×`EOR`) with two pointer-save/restore instructions (`STR.W` + `STR`): 24 theoretical cycles per call × 14 rounds = 336 cycles total.

https://doi.org/10.31436/iiumej.v27i2.4405

SubBytes: The Boyar–Peralta depth-16 Boolean circuit comprises 146 instructions (81 `EOR`, 32 `AND`, 16 `STR`, and 17 `LDR`), yielding 179 theoretical cycles per call. The measured value of 269 cycles gives a stall factor of 1.50 ×, attributable to frequent read-after-write hazards in the chained Boolean operations and load-use stalls on the stack-spilled intermediates.

ShiftRows: Each `shiftrows_one` invocation consists of 12 instructions (4 ×`UBFX`, 2 ×`ORR`, 2 ×`BFI` for rows 1–2, plus 2 ×`UBFX`, 2 ×`BFI` for the row-3 nibble swap) whose data dependencies are resolved by result forwarding on the Cortex-M4, giving 96 theoretical cycles for 8 registers per call. The measured value at `-O3` is 98 cycles, yielding a stall factor of 1.02 ×, essentially pipeline-stall-free. This confirms that the `UBFX/BFI` routing approach is not only correct but also architecturally efficient.

MixColumns: The Käsper–Schwabe circuit uses 27 `EOR/ROR` instructions (27 theoretical cycles), yet measures 82 cycles, a stall factor of 3.04 ×. This overhead is caused by the chained `EOR` sequence: many instructions read registers written 1–2 instructions earlier, causing systematic 1–2 cycle stalls on each dependency.

Unpacking: The reverse transposition comprises 12 swpmv invocations and mask-setup (`MOVW`, `MOVT`, 2 ×`EOR`), plus the function epilogue (`STMIA{8}`, `ADD.W`, `POP{14}`), yielding 77 theoretical cycles.

5.4.3. Summary

As summarized in Table 8, the theoretical minimum is 4,694 cycles; the measured result is 6,199 cycles, giving an overall pipeline efficiency of 4,694/6,199 ≈ 76%. Two findings are architecturally significant. First, ShiftRows achieves a stall factor of 1.02 ×, near-theoretical performance, because data dependencies are resolved by result forwarding on the Cortex-M4, causing no effective pipeline stalls. This result validates the explicit `UBFX/BFI` routing strategy as both functionally correct and pipeline-friendly. Second, MixColumns carries the highest stall penalty (3.04×) despite using only simple `EOR/ROR` instructions, because the Käsper–Schwabe diffusion circuit is inherently sequential with chained data dependencies. This suggests that reordering the MixColumns instruction stream to interleave independent operations across columns represents the most promising target for future micro-optimization.

Table 8. Theoretical instruction-level cost analysis for the macro-inlined assembly variant at `-O2`.

Component	Instr./call	Theory cyc/call	Calls	Total (theory)	Stall factor
Packing	57	80	1	80	≈1.1×
AddRoundKey	13	24	14	336	≈1.3×
SubBytes	146	179	14	2,506	≈1.5×
ShiftRows	96	96	14	1,344	1.02×
MixColumns	27	27	13	351	3.04×
Unpacking	56	77	1	77	≈1.6×
Total				**4,694**	**1.32×**

5.5. Code Size and Memory

The function-call assembly variant requires only 658 bytes of flash for the ASM cipher core with zero RAM overhead, while the macro-inlined variant requires 15,898 bytes due to full round-function expansion across all 14 rounds. The constant-time implementations require no data RAM beyond the stack and round keys. All implementations share 480 bytes of round key storage (15 keys × 32 bytes).

Table 9. Memory requirements for Rijndael-256 implementations on STM32F407 (`-O2`).

Implementation	Flash (B)	RAM (B)	Stack (B)	Round Keys (B)
Full ASM (macro)	15,898	0	112	480
Full ASM (func)	658	0	112	480
BS C (semi-f.)	2,204	0	96	480
T-table 1T	1,528	1,024	64	480
T-table 4T	5,680	4,096	64	480
Fixsliced HC	3,816	0	96	480

5.6. Constant-Time Verification

5.6.1. Methodology and Experimental Setup

To empirically verify the constant-time execution of the proposed bitsliced Rijndael-256 implementation, we adopted the DUDECT methodology proposed by Reparaz *et al.* [10]. Unlike static code analysis, DUDECT performs a leakage assessment using Welch's t-test by measuring actual execution times on the target hardware. The assessment evaluates whether the execution duration exhibits a statistically significant dependency on the processed input data. DUDECT therefore provides empirical evidence under the tested hardware and testbench conditions, not a formal proof of constant-time behavior.

The evaluation was conducted on an ARM Cortex-M4 microcontroller. Cycle-accurate measurements used the Data Watchpoint and Trace (DWT) cycle counter native to the Cortex-M architecture. The testing procedure was divided into two distinct classes: fixed data (Class 0) and randomized data (Class 1). The t-statistic was continuously computed using Welford's online algorithm for running variance, enabling the collection of 10^6 samples without requiring $O(N)$ memory on the target device. A $|t|$-statistic value exceeding 4.5 indicates detectable timing leakage with a confidence level greater than 99.999%.

A critical aspect of hardware-level timing analysis is isolating software-induced noise and architectural artifacts. During our initial setup, the standard C library pseudorandom number generator (PRNG) was found inadequate because its internal state management introduced artificial memory-alignment stalls. Consequently, we integrated the microcontroller's hardware true random number generator (TRNG) to supply the testbench with high-entropy random values, ensuring unbiased class distribution and input generation. Furthermore, to mitigate pipeline state carry-over caused by the differing data preparation routines of Class 0 and Class 1, Data Synchronization Barrier (`DSB`) and Instruction Synchronization Barrier (`ISB`) instructions were injected immediately preceding the DWT cycle capture. This guarantees a uniform micro-architectural state prior to execution [18].

5.6.2. Results and Discussion

The implementation was compiled using the GCC toolchain and evaluated across all primary optimization levels (`-O0`, `-O1`, `-O2`, `-O3`, and `-Os`). For each configuration, we collected $N = 10^6$ samples. The progression of the experimental results, summarized in Table 10, highlights the crucial role of environmental sanitization in micro-architectural timing evaluations.

Initial baseline testing with the software PRNG reported false-positive leakages at the -O1 optimization level ($t = 0.09$). This anomaly was traced to memory alignment discrepancies introduced by the compiler when handling the PRNG's static variables alongside the testbench buffers, resulting in an artificial penalty of approximately 2 cycles. Replacing the PRNG with the hardware TRNG eliminated this artifact, dropping the t-statistic to 0.11 and confirming the absence of algorithmic leakage at `-O1`.

https://doi.org/10.31436/iiumej.v27i2.4405

Table 10. DUDECT Evaluation Results for Full ASM (macro) Variant Across Compiler Optimization Levels ($N = 10^6$ Samples).

Opt. Level	Testbench Environment	Mean Class 0	Mean Class 1	t-Statistic / Verdict
`-O0`	Baseline (PRNG)	6208.64	6208.61	1.1173 / PASS
`-O0`	TRNG Only (No Flush)	6208.67	6208.68	0.3151 / PASS
`-O0`	Final (TRNG + Pipeline Flush)	6209.61	6209.61	0.0285 / PASS
`-O1`	Baseline (PRNG)	6202.15	6204.11	90.09 / FAIL (False Positive)
`-O1`	TRNG Only (No Flush)	6205.08	6205.08	0.1054 / PASS
`-O1`	Final (TRNG + Pipeline Flush)	6202.07	6202.07	0.1912 / PASS
`-O2`	Baseline (PRNG)	6197.64	6197.64	0.3425 / PASS
`-O2`	TRNG Only (No Flush)	6198.72	6199.75	57.45 / FAIL (1-Cycle Stall)
`-O2`	Final (TRNG + Pipeline Flush)	6198.71	6198.71	0.2385 / PASS
`-O3`	Baseline (PRNG)	6197.83	6197.84	0.8650 / PASS
`-O3`	TRNG Only (No Flush)	6198.64	6199.63	59.16 / FAIL (1-Cycle Stall)
`-O3`	Final (TRNG + Pipeline Flush)	6198.73	6198.77	2.0229 / PASS
`-Os`	Baseline (PRNG)	6197.88	6197.85	1.8792 / PASS
`-Os`	TRNG Only (No Flush)	6197.93	6197.94	0.5411 / PASS
`-Os`	Final (TRNG + Pipeline Flush)	6197.86	6197.87	0.3785 / PASS

However, isolating the input generation with the TRNG revealed a secondary micro-architectural artifact at higher optimization tiers (`-O2` and `-O3`). As shown in the intermediate TRNG tests, the evaluation yielded t-statistics exceeding 55. A detailed trace of the cycle counts showed a microscopic, highly stable difference of exactly 1 cycle between the two classes (e.g., 6198.72 vs. 6199.75 cycles at `-O2`).

This 1-cycle discrepancy is a classic manifestation of branch predictor and pipeline state leakage. The testbench prepares Class 0 using a highly predictable `memset` operation, whereas Class 1 requires iterative polling of the TRNG peripheral. At `-O2` and `-O3`, the aggressive compiler loop unrolling and branch generation for the TRNG polling left the Cortex-M4's pipeline in an asymmetric state upon entering the encryption function. The 1-cycle delay was due to the processor flushing its pipeline after a loop-termination misprediction.

By implementing the `DSB` and `ISB` pipeline flushing mechanism prior to the measurement block, the processor state was successfully synchronized. As demonstrated in Table 10, the t-statistics for all optimization levels, including `-O2` ($t = 0.23$) and -O3 ($t = 2.02$), fell well below the 4.5 threshold. The execution cycles became well-aligned across both classes, providing strong statistical evidence that the macro-inlined assembly variant exhibits no detectable data-dependent timing variation across all tested compiler optimization levels.

Table 11. DUDECT Evaluation Results for Full ASM (func) Variant ($N = 10^6$ Samples, TRNG + Pipeline Flush).

Opt. Level	Mean Class 0	Mean Class 1	t-Statistic / Verdict
`-O0`	6534.72	6534.68	1.7204 / PASS
`-O1`	6527.14	6527.10	1.7555 / PASS
`-O2`	6523.77	6523.73	2.1849 / PASS
`-O3`	6523.79	6523.80	0.7758 / PASS
`-Os`	6522.90	6522.90	0.3333 / PASS

The function-call variant (Full ASM (func)) shares an identical assembly core with the macro-inlined variant; the sole architectural difference is the addition of data-independent function-call overhead (BL/BX and PUSH/POP instructions), which, by definition, introduces

https://doi.org/10.31436/iiumej.v27i2.4405

no secret-dependent timing variation. Full DUDECT evaluation of this variant with $N = 10^6$ samples using the same TRNG and pipeline-flush methodology is presented in Table 11. All optimization levels yield t-statistics well below the threshold of 4.5, with a maximum t-statistic of 2.18 at `-O2`, confirming that the constant-time property is preserved across both variants.

6. COMPARISON WITH AES-128 AND AES-256

Table 12 summarizes the constant-time performance comparison. First, we compare Rijndael-256 with AES-128 using the public fully-fixsliced AES implementation of Adomnicai and Peyrin [9], re-measured on the same STM32F407VG platform and under the same DWT-based timing protocol used for our Rijndael-256 implementation. Under this same-platform measurement, AES-128 achieves 105.1 cycles/byte. Therefore, the rolled Rijndael-256 variant yields $203.7/105.1 = 1.94\times$, while the unrolled variant yields $193.7/105.1 = 1.84\times$. The originally reported AES-128 figure in [9] is 80.0~cycles/byte; using that published value instead would give ratios of $203.7/80 = 2.55\times$ for the rolled variant and $193.7/80 = 2.42\times$ for the unrolled variant. We therefore use the same-platform reproduced value as the primary normalization point, while reporting the published-reference ratios for transparency.

The expected scaling from AES-128 to Rijndael-256 is affected by two factors. Rijndael-256 uses 14 rounds instead of 10 (factor $1.4\times$), while the fixsliced AES-128 implementation processes two 128-bit blocks in parallel ($2\times$ throughput advantage). A simple scaling estimate suggests $1.4\times 2.0 = 2.8\times$. Both the same-platform ratio of $(1.84\times - 1.94\times)$ and the published-reference ratio of $(2.42\times - 2.55\times)$ are below this estimate, indicating that the proposed semi-fused Rijndael-256 strategy is efficient relative to the naive scaling expectation, despite the wider state and irregular ShiftRows structure.

Table 12. Comparison of constant-time performance on STM32F407VG at 168 MHz. All cycle counts include packing and unpacking.

Cipher	Key (bit)	Blocks	Rounds	Cycles	Cycles/byte
AES-128†	128	2	10	3,362	105.1
AES-256†	256	2	14	4,590	143.4
Rijndael-256 (rolled)	256	1	14	6,518	203.7
Rijndael-256 (unrolled)	256	1	14	6,199	193.7

† AES-128 and AES-256 were measured using the public fully-fixsliced implementation of Adomnicai and Peyrin [9] on the same STM32F407VG platform and under the same interrupt-disabled DWT measurement protocol. Under this same-platform setup, AES-128 measured 105.1 cycles/byte. The originally reported AES-128 figure in [9] is 80.0 cycles/byte; we use the reproduced value on the same platform for the primary comparison.

Second, we compare Rijndael-256 with AES-256. This comparison is structurally cleaner because both ciphers use a 256-bit key, perform 14 rounds, and process 32 bytes per call (AES-256 through two parallel 128-bit blocks, and Rijndael-256 through one 256-bit block). Using the same platform AES-256 measurement of 143.4 cycles/byte, the rolled Rijndael-256 variant yields $203.7/143.4 = 1.42\times$, while the unrolled variant yields $193.7/143.4 = 1.35\times$. This ratio more directly isolates the cost of Rijndael-256's wider state ($N_b = 8$) and irregular ShiftRows offsets relative to AES-256 ($N_b = 4$), with key size and round count held constant.

For context, Saarinen reports a $2.66\times$ instruction count ratio for Rijndael-256 versus AES-256 on RISC-V Vector hardware [6]. That metric targets a different architecture and measures instructions rather than cycles; our $1.42\times$ cycle ratio on Cortex-M4 reflects a different trade-off space but confirms that Rijndael-256 imposes a moderate, not prohibitive, overhead relative to AES-256.

https://doi.org/10.31436/iiumej.v27i2.4405

7. DISCUSSION

7.1. Fixslicing Optimality is Cipher-Specific

Our central finding is that fixslicing is suboptimal for Rijndael-256 while being optimal for AES-128, and this has implications for the implementation of other Rijndael variants. For AES-128 ($N_b = 4$), the column field width $g = 8$ bits ensures that all column shifts are byte-aligned, allowing the barrel shifter to absorb them at zero instruction cost. For Rijndael-256 ($N_b = 8$), $g = 4$ bits makes column shifts sub-byte, requiring per-row `UBFX/BFI` manipulation. This distinction is a property of the interaction between the cipher's block size N_b and the target register width W, not an inherent limitation of the fixslicing technique itself.

7.2. The 1-Table Advantage on ARM

The 20% performance advantage of 1-table over 4-table on Cortex-M4 at `-O2` is architecturally significant. On RISC-V RV32I, the additional cycles for software rotation would reduce or eliminate this advantage. Both this result and Saarinen's findings [6] demonstrate that the relative cost of implementation strategies depends heavily on available ISA primitives.

7.3. Compiler Sensitivity

A 6% performance loss occurs when transitioning from `-O2` to `-O3` in the bitsliced implementation, primarily due to code size expansion (from 2,204 to 3,732 bytes) that exceeding the instruction cache working set. This results in increased cache misses, negating the advantages of aggressive optimization. Conversely, the full assembly variants remain unaffected due to their compact code size.

7.4. Macro Inlining: Throughput vs. Code Size Trade-off

The two assembly variants expose a fundamental trade-off between throughput and flash footprint that is particularly consequential in constrained IoT deployments. The macro-inlined variant (Full ASM (macro)) expands all 14 round functions, each comprising `ark_sbox`, `shiftrows_rj256`, and `mixcolumns_rj256`, at their respective call sites, eliminating `BL/BX` branching and `PUSH/POP` register save/restore overhead. This yields a 319-cycle reduction per encryption (6,518 to 6,199 cycles, or 4.9%), at the cost of 15,240 additional bytes of flash (658 to 15,898 bytes, a 24.2 × increase).

To contextualize this trade-off, a device with 16 KB of flash can accommodate the function-call variant alongside a minimal protocol stack, whereas the macro-inlined variant alone consumes nearly the entire flash of a 16 KB device and is therefore only suitable for targets with at least 32 KB of flash. For high-throughput applications on flash-unconstrained devices, the macro variant offers a measurable throughput advantage. For typical IoT sensor nodes, the function-call variant is the practical choice.

This disproportion, large code size increase for a modest cycle reduction, arises because the Rijndael-256 round function is substantially larger than its AES-128 counterpart (4 × 8 vs. 4 × 4 state). On AES-128, fixsliced macro inlining yields a more favorable ratio because the round function is smaller. This further illustrates that the implementation strategy must be tailored to the specific cipher and target constraints.

https://doi.org/10.31436/iiumej.v27i2.4405

7.5. Practical Implications for IoT Deployment

At 203.7 cycles/byte, the function-call assembly variant achieves approximately 825 KB/s at 168 MHz, which is sufficient for typical IoT workloads; a 256 KB firmware image would encrypt in approximately 0.318 seconds. Using the STM32F407 active-mode current of approximately 87 mA at 168 MHz and a 3.3 V supply, the estimated active energy is about 11.1 µJ per 32-byte block, 0.348 µJ/byte, or 91 mJ for a 256 KiB firmware image. In CTR mode, a split pack/encrypt-loop/unpack implementation could reduce the per-block cost by amortizing packing and unpacking across consecutive blocks, though such restructuring is beyond the scope of this work.

For DTLS-secured sensor data with typical payloads of 64–256 bytes, the encryption cost is 2–8 Rijndael-256 block operations, completing in approximately 78–310 µs on the function-call assembly variant and consuming roughly 22–89 µJ of active energy under the estimate above. The 658-byte ASM core footprint of the function-call variant allows the cipher to coexist with protocol stacks even on devices with as little as 16 KB of flash. The macro-inlined variant (15,898 bytes) is more appropriate for targets with larger flash budgets where maximum throughput is the priority.

For deployments where timing-side-channel resistance is not required (e.g., isolated sensor nodes without physical access), the T-table 1T variant at 101.6 cycles/byte offers higher throughput at 1 KB RAM cost, but it should not be used when cache-timing leakage is within the threat model.

7.6. Limitations and Future Work

This work focuses on the encryption direction, which suffices for the dominant use cases of Rijndael-256 in practice: CTR, GCM, and CCM modes require only the block cipher's forward permutation. The implementation is single-block and software-only on ARM Cortex-M4. Decryption requires InvSubBytes, InvShiftRows, and InvMixColumns; the bitsliced InvMixColumns can be decomposed as $\mathrm{InvMC}(x) = \mathrm{MC}(M_{extra}(x))$ [11], enabling the reuse of the existing MixColumns circuit with a preprocessing step. A complete decrypt implementation is left as future work.

The constant-time property established here addresses the timing side-channel threat model. Power and electromagnetic side-channel attacks constitute a distinct threat class requiring masking countermeasures, and fault attacks require separate redundancy or detection mechanisms. The bitsliced representation is structurally amenable to first-order Boolean masking; the linear layers are maskable at negligible cost, and the Boolean-domain operations eliminate the data-dependent memory accesses that complicate masking of T-table implementations. A complete first-order masked implementation would require a DOM-based masked S-Box circuit and a masked key schedule. This remains future work rather than a claim of present resistance.

Extending this analysis to Rijndael-192 and to other 32-bit architectures (RISC-V RV32IM, Cortex-M33) would further validate our findings. Direct power-trace measurements would complement the cycle-count analysis and the back-of-the-envelope energy estimate reported in Section 7.5.

8. CONCLUSION

We have presented, to the best of our knowledge, the first constant-time bitsliced implementation of Rijndael-256 on the ARM Cortex-M4. This work addresses a practical need arising from NIST's ongoing consideration of Rijndael-256 for standardization (SP 800-197).

Our principal technical contribution is a quantitative analysis demonstrating that fixslicing becomes suboptimal for Rijndael-256 on this platform. The ShiftRows period $T = 8$ for $N_b = 8$ requires eight MixColumns compensation variants, costing on average $3.00\times$ more than explicit ShiftRows, making the semi-fused approach 17% faster, inverting the relationship established for AES-128.

The macro-inlined assembly variant achieves 193.7 cycles/byte, and the function-call variant achieves 203.7 cycles/byte, both with zero RAM overhead. The function-call variant requires only 658 bytes of flash, while the macro-inlined variant trades 15,898 bytes of flash for a 319-cycle savings per block. The constant-time overhead of approximately 1.91× relative to the T-table baseline (6,199 vs. 3,252 cycles) represents a practical cost for timing side-channel resistance on constrained devices.

These results demonstrate that the optimal bitslicing strategy depends on the interaction between cipher-specific ShiftRows offsets, block size, and the target architecture's instruction set. As Rijndael-256 moves toward potential standardization, implementers should evaluate rather than assume that techniques optimal for AES-128 will transfer to other Rijndael variants. The source code and measurement artifacts are available from the corresponding author upon reasonable request and will be archived publicly upon acceptance.

ACKNOWLEDGEMENT

The work of Andriani Adi Lestari was supported by Lembaga Pengelola Dana Pendidikan (LPDP), Ministry of Finance of the Republic of Indonesia, under Contract SKPB-4400/LPDP/LPDP.3/2023.

REFERENCES

[1] National Institute of Standards and Technology. (2023) Advanced Encryption Standard (AES). Federal Information Processing Standards Publication (FIPS) 197-upd1. doi:10.6028/nist.fips.197-upd1

[2] Dworkin MJ. (2007) Recommendation for Block Cipher Modes of Operation: Galois/Counter Mode (GCM) and GMAC. NIST Special Publication 800-38D. doi:10.6028/NIST.SP.800-38D

[3] McGrew DA, Viega J. (2005) The Security and Performance of the Galois/Counter Mode (GCM) of Operation. In: Progress in Cryptology -- INDOCRYPT 2004. Springer Berlin Heidelberg: 343-355. doi:10.1007/978-3-540-30556-9_27

[4] Avanzi R, Chakraborti A, Chakraborty B, List E. (2025) The Large Block Cipher Vistrutah. IACR Transactions on Symmetric Cryptology. 2025(3): 1-150. doi:10.46586/tosc.v2025.i3.1-150

[5] National Institute of Standards and Technology. (2024) Initial Pre-Draft Call for Comments on SP 800-197. NIST Cybersecurity and Privacy Reference Tool. URL: https://csrc.nist.gov/pubs/sp/800/197/iprd

[6] Saarinen MJO. (2025) Brief Comments on Rijndael-256 and the Standard RISC-V Cryptography Extensions. Cryptology ePrint Archive, Paper 2025/1198. URL: https://eprint.iacr.org/2025/1198

[7] Mordor Intelligence. (2025) IoT Microcontroller Market Size & Share Analysis–Growth Trends and Forecast (2025–2030). Online; accessed Apr. 2026. URL: https://www.mordorintelligence.com/industry-reports/iot-microcontroller-market

[8] IoT Analytics. (2025) IoT Microcontrollers Market Report 2025–2030. Market Report, Oct. 2025. URL: https://iot-analytics.com/product/iot-microcontroller-market-report-2025-2030/

[9] Adomnicai A, Peyrin T. (2021) Fixslicing AES-like Ciphers: New Bitsliced AES Speed Records on ARM-Cortex M and RISC-V. IACR Transactions on Cryptographic Hardware and Embedded Systems. 2021(1): 402-425. doi:10.46586/tches.v2021.i1.402-425

[10] Reparaz O, Balasch J, Verbauwhede I. (2017) Dude, is my code constant time? In: Design, Automation & Test in Europe Conference & Exhibition (DATE), 2017: 1697-1702. doi:10.23919/DATE.2017.7927267

[11] Daemen J, Rijmen V. (2002) The Design of Rijndael: AES — The Advanced Encryption Standard. Springer-Verlag, Berlin, Germany.

[12] Biham E. (1997) A Fast New DES Implementation in Software. In: Fast Software Encryption, Lecture Notes in Computer Science, vol. 1267. Springer Berlin Heidelberg: 260-272. doi:10.1007/BFb0052352

[13] Boyar J, Peralta R. (2012) A Small Depth-16 Circuit for the AES S-Box. In: Information Security and Privacy Research, IFIP Advances in Information and Communication Technology, vol. 376. Springer Berlin Heidelberg: 287-298. doi:10.1007/978-3-642-30436-1_24

[14] Schwabe P, Stoffelen K. (2017) All the AES You Need on Cortex-M3 and M4. In: Selected Areas in Cryptography -- SAC 2016, Lecture Notes in Computer Science, vol. 10532. Springer: 180-194. doi:10.1007/978-3-319-69453-5_10

[15] Bernstein DJ. (2005) Cache-Timing Attacks on AES. Technical report. URL: https://cr.yp.to/antiforgery/cachetiming-20050414.pdf

[16] Almeida JB, Barbosa M, Barthe G, Dupressoir F, Emmi M. (2016) Verifying Constant-Time Implementations. In: 25th USENIX Security Symposium. USENIX Association: 53-70.

[17] Adomnicai A, Najm Z, Peyrin T. (2020) Fixslicing: A New GIFT Representation: Fast Constant-Time Implementations of GIFT and GIFT-COFB on ARM Cortex-M. IACR Transactions on Cryptographic Hardware and Embedded Systems. 2020(3): 402-427. doi:10.13154/tches.v2020.i3.402-427

[18] ARM Limited. (2010) Cortex-M4 Technical Reference Manual. DDI 0439B. URL: https://developer.arm.com/documentation/ddi0439/b

[19] STMicroelectronics. (2024) RM0090 Reference Manual: STM32F405/415, STM32F407/417, STM32F427/437 and STM32F429/439 Advanced Arm-based 32-bit MCUs. RM0090, Rev. 21.

[20] Käsper E, Schwabe P. (2009) Faster and Timing-Attack Resistant AES-GCM. In: Cryptographic Hardware and Embedded Systems -- CHES 2009, Lecture Notes in Computer Science, vol. 5747. Springer Berlin Heidelberg: 1-17. doi:10.1007/978-3-642-04138-9_1

https://doi.org/10.31436/iiumej.v27i2.4072

Scalable Multi-Objective Optimization for Facility Location Using a Metaheuristic Technique

THANATHORN KAROT, CHOOSAK PORNSING*

Department of Industrial Engineering and Management, Faculty of Engineering and Industrial Technology, Silpakorn University, Nakhon Pathom 73000, Thailand

**Corresponding author: pornsing_c@su.ac.th*

(Received: 30 November 2025; Accepted: 21 February 2026; Published online: 10 May 2026)

ABSTRACT: Rapid growth in e-commerce and service expectations forces large-scale logistics networks to balance facility operating cost, service coverage, and delivery equity simultaneously. This study aims to develop a scalable discrete multi-objective particle swarm optimization (MOPSO-FLP) framework for constraint-rich facility location problems, where traditional exact approaches become impractical. The proposed method integrates discrete encoding, feasibility repair, adaptive parameter control, and an archive-based leader selection strategy, and it is evaluated using benchmark instances and a national postal logistics network case. Across 30 independent runs, MOPSO-FLP achieves superior convergence and diversity compared with representative multi-objective metaheuristics (e.g., NSGA-II and related baselines), and the deployment validation yields a Pareto set of 47 non-dominated solutions that clearly exposes cost–coverage–equity trade-offs. Overall, the results demonstrate that the proposed framework provides decision makers with interpreted alternatives and actionable policies for large-scale logistics planning.

ABSTRAK: Pertumbuhan pesat dalam e-dagang serta peningkatan jangkaan perkhidmatan memaksa rangkaian logistik berskala besar untuk mengimbangi kos operasi fasiliti, liputan perkhidmatan, dan keadilan penghantaran secara serentak. Kajian ini bertujuan membangunkan satu kerangka pengoptimuman zarah berbilang objektif diskret yang berskala, iaitu Multi-Objective Particle Swarm Optimization bagi masalah penentuan lokasi fasiliti (MOPSO-FLP), khusus untuk persekitaran berkekangan tinggi di mana pendekatan tepat tradisional menjadi tidak praktikal. Kaedah yang dicadangkan mengintegrasikan pengekodan diskret, mekanisme pembaikan kebolehlaksanaan, kawalan parameter adaptif, serta strategi pemilihan pemimpin berasaskan arkib, dan dinilai menggunakan set data penanda aras serta kajian kes rangkaian logistik pos kebangsaan. Merentasi 30 ulangan bebas, MOPSO-FLP menunjukkan penumpuan dan kepelbagaian penyelesaian yang lebih baik berbanding metaheuristik berbilang objektif sedia ada seperti NSGA-II dan kaedah asas berkaitan. Pengesahan pelaksanaan menghasilkan satu set Pareto yang mengandungi 47 penyelesaian tidak didominasi, yang secara jelas menggambarkan kompromi antara kos, liputan, dan keadilan. Secara keseluruhan, dapatan kajian ini menunjukkan bahawa kerangka yang dicadangkan mampu menyediakan alternatif yang boleh ditafsir serta dasar yang boleh dilaksanakan oleh pembuat keputusan dalam perancangan logistik berskala besar.

KEYWORDS: *Facility location, Logistics systems, Multi-objective Optimization, Particle swarm optimization, Metaheuristics*

1. INTRODUCTION

Rapid changes in demand patterns, e-commerce growth, and service expectations have complicated the design and management of logistics networks. A central strategic question in

https://doi.org/10.31436/iiumej.v27i2.4072

such systems is where to locate facilities, such as warehouses, sorting centers, or service hubs, to consistently balance cost, service accessibility, and responsiveness. Facility location problems (FLPs) provide a mathematical framework for this decision, but real applications are rarely driven by a single objective. Instead, planners typically face several criteria that pull in different directions, for example, low cost versus high coverage or short travel times.

Traditional FLP models and solution methods have often been built around a single objective, most commonly cost minimization. While these formulations have contributed to theory and practice, they can mask important trade-offs and fail to account for issues such as service equity, environmental concerns, or robustness to disruptions [1, 2]. In many current applications, decision makers need to see not one "optimal" configuration but a spectrum of alternatives that trade cost for performance in different ways.

To support this need, multi-objective optimization (MOO) approaches that approximate the Pareto front have attracted increasing attention [3, 4]. In a Pareto-based view, a solution is considered efficient if no objective can be improved without worsening at least one other objective. Approximating such a front is particularly demanding when the underlying decision space is discrete and highly constrained, which is typical in realistic FLPs with capacity limits, service radius requirements, and policy rules.

Exact methods can, in principle, handle multi-objective models, but their computational effort tends to grow rapidly with problem size and constraint complexity. For national-scale or highly detailed regional systems, purely exact algorithms become difficult to use as routine planning tools. In contrast, metaheuristic algorithms seek high-quality solutions with more modest computational requirements, trading mathematical guarantees for scalability and flexibility.

Particle Swarm Optimization (PSO) is one such metaheuristic that has seen widespread use in engineering optimization because of its simple structure and few parameters [5, 6]. Its multi-objective variants (MOPSO) extend the original idea by maintaining an archive of non-dominated solutions and using leaders drawn from this archive to guide the swarm. Although a number of MOPSO approaches have been proposed, their direct application to discrete, constraint-rich FLPs remains limited, especially when several real-world considerations, such as coverage, equity, and maximum travel time, must be accounted for simultaneously [7].

This work aims to narrow that gap by developing a discrete MOPSO framework tailored specifically to multi-objective FLPs. The algorithm uses a binary encoding for facility opening decisions, while assignments and coverage are determined deterministically from those decisions. In addition to the basic swarm update rules, the algorithm uses adaptive parameter settings, an archive from which leaders are selected, and simple repair and local-search moves to satisfy constraints while preserving sufficient diversity in the swarm. The main contributions of the study are: 1) A discrete multi-objective facility location formulation is proposed that integrates elements of the p-median, maximal covering, and p-center models into a unified framework capturing cost, service coverage, and spatial equity, 2) A MOPSO-based solution procedure is developed, featuring problem-oriented encoding, feasibility-repair mechanisms, and adaptive parameter control to address large-scale and highly constrained search spaces, 3) The algorithm is examined through an extensive set of experiments on benchmark and synthetically generated instances of various sizes, with performance evaluated by several quality indicators and supported by statistical analyses of robustness and scalability, and 4) The framework is further demonstrated in the context of redesigning a national postal logistics network, where the obtained Pareto solutions are used to illustrate how alternative network configurations can inform managerial decision-making.

https://doi.org/10.31436/iiumej.v27i2.4072

Novelty and contributions: This work advances the multi-objective facility location literature in three ways. (i) It proposes a discrete MOPSO formulation tailored to facility location decisions with practical constraints, including discrete site selection and operational feasibility handling. (ii) It integrates adaptive parameter control, feasibility repair, and archive-based leader selection to improve scalability and maintain diversity in high-dimensional, discrete search spaces. (iii) It establishes a consistent evaluation protocol with benchmark comparisons and statistical validation and demonstrates managerial interpretability through a national postal logistics network redesign case.

The paper is divided into five main sections. Section 2 presents a focused review of facility location and multi-objective metaheuristics. Section 3 describes the mathematical formulation and the proposed MOPSO-FLP algorithmic design. Section 4 reports computational experiments, sensitivity analyses, and validation of the national postal network deployment. Section 5 concludes the paper and outlines directions for future research.

2. PREVIOUS WORKS

2.1. Facility Location Problems

Facility location problems have been a core topic in location science and logistics for many decades. The earliest formulations were typically deterministic and relied on a single performance measure. Classic examples include the p-median model, which seeks to minimize the total demand-weighted distance or transportation cost, and the p-center model, which focuses on reducing the worst (maximum) distance between demand points and open facilities. A third foundational model is maximal covering, where the aim is to maximize the amount of demand served within a given service distance or time limit. These three formulations have shaped much of the subsequent work in the area and still serve as reference points for more elaborate models.

Over time, researchers have moved from these basic formulations to more elaborate FLPs that more closely mirror real planning contexts. Contemporary models often combine several objectives, introduce capacity constraints at facilities, and incorporate uncertainty in demand levels or travel times. Issues of fairness and equity in access to services have also become more visible, especially in public-sector and humanitarian applications. Fugaro and Sgalambro [8] review this evolution and note a growing interest in multi-criteria formulations that reflect social and environmental concerns alongside traditional economic objectives.

2.1. Metaheuristics in FLPs

As FLP formulations become more complex, metaheuristics have emerged as practical tools to search large solution spaces. Genetic Algorithms (GAs), Ant Colony Optimization (ACO), PSO, and other nature-inspired techniques have been successfully employed in various location and layout settings.

Koosha et al. [10], for instance, applied a GA to an unequal-area facility layout problem, balancing safety and cost under geometric constraints. ACO-based methods have been effective in routing and distribution problems on realistic road networks [11]. PSO, with its swarm-based search, has been adapted for tasks such as selecting locations for logistics centers or emergency facilities, particularly when decisions must be updated in response to changing conditions [12].

Within multi-objective optimization, MOPSO extends PSO by maintaining a collection of non-dominated solutions and using archive-based leaders to guide particles toward different

regions of the Pareto front. Work by Makhadmeh et al. [13] and Kaveh et al. [14] demonstrates that carefully designed MOPSO variants can deliver competitive hypervolume and diversity metrics in industrial and construction applications, highlighting the potential of PSO-based methods for problems with conflicting criteria.

2.3. Multi-Objective Particle Swarm Optimization (MOPSO)

MOPSO variants differ mainly in how they manage the external archive, how leaders are selected, and how diversity is preserved along the Pareto front. Common strategies include grid-based mechanisms, crowding distance measures, and restricted tournament selection. These design choices influence both convergence speed and the spread of solutions across the objective space.

In most studies, the quality of a multi-objective algorithm is judged using the hypervolume indicator, which quantifies the region of the objective space dominated by the obtained solutions, together with measures such as inverted generational distance and spacing. Empirical comparisons indicate that carefully tuned MOPSO variants can achieve high hypervolume values and compete on a wide range of engineering benchmarks [13].

MOPSO has also been applied to problems well beyond traditional logistics and scheduling. Recent examples include renewable energy planning and microgrid design, where economic, technical, and, sometimes, environmental objectives must be considered simultaneously. These applications illustrate that MOPSO can handle heterogeneous objectives and large, complex search spaces, though problem-specific modifications are often needed to achieve good performance in each new domain.

2.4. Thematic Synthesis: Methodological Gaps and Future Directions

Although considerable progress has been made in MOPSO and related approaches, several open issues remain in the literature. A large share of existing MOPSO variants has been developed with continuous decision spaces in mind, using real-valued positions and velocities. When these schemes are transferred to discrete facility location settings without further adaptation, they often generate infeasible solutions or converge to configurations that are clearly suboptimal [15].

Another concern relates to how algorithms are benchmarked. Different authors rely on different instance sets, performance indicators, and stopping rules, making it difficult to compare results across studies or to reproduce reported findings consistently [16]. Hybrid strategies that combine metaheuristics with decomposition, column generation, or other exact techniques have been proposed as a possible remedy, but their role in multi-objective FLPs remains only partially explored [17].

In terms of application domains, most published work focuses on logistics, emergency response, or healthcare. Areas such as climate-resilient infrastructure, autonomous or intelligent transportation systems, and equity-oriented urban planning have received much less attention so far. Uncertainty is frequently addressed through a small set of static scenarios, rather than through richer stochastic or dynamic models, and ethical dimensions, such as equitable access to services or the explicit involvement of stakeholders in defining objectives, are only starting to appear systematically [8].

The present study is motivated by these observations and proposes a discrete MOPSO framework tailored to the combinatorial nature of FLPs, uses standardized benchmarks and common performance indicators, and incorporates equity-related objectives directly into the formulation, allowing issues of fairness to be examined alongside cost and efficiency.

https://doi.org/10.31436/iiumej.v27i2.4072

2.5. Positioning the Current Study

In light of the above discussion, the proposed MOPSO-FLP framework is a discrete, multi-objective metaheuristic specifically tailored to facility location problems under realistic constraints and explicit equity considerations. Facilities are encoded as a binary vector indicating which sites are open, while assignments and coverage levels are determined by simple deterministic rules. This separation keeps feasibility checks relatively straightforward and allows us to design repair moves and local search operators that act directly on the location decisions.

A second feature of the study is the emphasis on a consistent evaluation protocol. We test the algorithm on both benchmark and synthetically generated instances of varying sizes, use commonly used performance indicators, and perform statistical analyses to compare results across algorithms. The intention is to make the assessment as transparent and reproducible as possible. Finally, the framework is exercised on the redesign of a national postal logistics network, which serves as a concrete illustration of how the method can be used in practice and how the resulting trade-off solutions may be relevant not only to postal services but also to related areas such as disaster logistics, environmentally oriented infrastructure planning, and equitable provision of public services.

3. METHOD

This section presents the methodological framework adopted for modeling and solving the multi-objective facility location problem. The approach combines a multi-objective facility location formulation with a discrete MOPSO algorithm and a systematic experimental design.

3.1. Problem Formulation

3.1.1. Multi-objective Facility Location Model

The considered FLP is defined on a bipartite graph $G = (I \cup J, E)$, where $I = \{1,2,\dots,n\}$ denotes demand nodes and $J = \{1,2,\dots,m\}$ denotes candidate facility sites. Each demand node $i \in I$ has demand w_i, and opening a facility at the site $j \in J$ incurs a fixed cost f_j. The distance between the demand point i and the site j is d_{ij}, which may represent Euclidean or network distance. Two global parameters are specified: the maximum number of facilities p and a service radius r.

The decision variables are:

- $y_j \in \{0,1\}$: 1 if a facility is opened at the site j, 0 otherwise;
- $x_{ij} \in \{0,1\}$: 1 if the demand node i is assigned to the facility j;
- $c_i \in \{0,1\}$: 1 if demand node i is considered covered within the radius r.

Three conflicting objectives are modeled:

1. Total cost (fixed plus transportation):

$$F_1 = \sum_{j\in J} f_j y_j + \sum_{i\in I}\sum_{j\in J} w_i d_{ij} x_{ij} \quad (1)$$

2. Weighted service coverage (to be maximized):

$$F_2 = \sum_{i\in I} w_i c_i \quad (2)$$

3. Maximum allocation distance (to be minimized):

https://doi.org/10.31436/iiumej.v27i2.4072

$$F_3 = \max_{i \in I, j \in J} \{d_{ij} x_{ij}\} \quad (3)$$

The model is subject to the following constraints:

$$\sum_{j \in J} x_{ij} = 1, \quad \forall i \in I \quad (4)$$

$$x_{ij} \leq y_j, \quad \forall i \in I, \forall j \in J \quad (5)$$

$$\sum_{j \in J} y_j \leq P \quad (6)$$

$$\sum_{j \in N} y_j \geq c_i, \quad \forall i \in I, where\ N_i = \{j \in J : d_{ij} \leq r\} \quad (7)$$

Equation (4) ensures that each demand node is assigned to exactly one facility, while Eq. (5) enforces that assignments can only be made to open facilities. The cardinality constraint in Eq. (6) limits the total number of facilities that may be opened, and Eq. (7) defines the coverage variable c_i in terms of the service radius r. Taken together, these equations capture the trade-offs between investment and operating cost, service coverage, and spatial equity

3.1.2. Problem Characteristics

The resulting optimization problem is combinatorial and NP-hard. The binary facility decisions induce a vast number of possible network configurations, and assignment and coverage variables are tightly linked to those decisions. The objectives are structurally conflicting: reducing the number of facilities typically lowers fixed costs but may deteriorate coverage and increase the maximum distance; expanding the network has the opposite effect. Because of these features, exact methods scale poorly to large instances, which motivates the use of an efficient multi-objective metaheuristic.

3.2. MOPSO-FLP Algorithm Design

The proposed MOPSO-FLP algorithm adapts PSO to the multi-objective and binary setting of FLPs. Each particle encodes a candidate set of open facilities through a binary vector $y \in \{0,1\}^m$. Given a facility vector, assignments x_{ij} and coverage variables c_i are derived by deterministic rules, for example, by assigning each demand node to its nearest open facility, which simplifies evaluation and ensures feasibility once y satisfies the facility limit.

Particle velocities are interpreted as probabilities of switching bits in y. A standard binary PSO update is used: for each component, the velocity combines inertia, cognitive, and social terms; a sigmoid function maps the velocity to a probability; and the bit is flipped with that probability. An external archive of non-dominated solutions is maintained, and leaders are drawn from this archive using a roulette mechanism weighted by a diversity indicator (e.g., crowding distance) to encourage exploration of different regions of the Pareto front.

Adaptive parameter control is employed: the inertia weight ω decreases from a high initial value to a lower final value over the iterations, gradually shifting the search from exploration to exploitation. The acceleration coefficients c_1 and c_2 can also be adjusted over time to balance individual learning and social influence. The main steps of the proposed MOPSO-FLP procedure are summarized in the flowchart shown in Fig. 1, which highlights the initialization phase, archive update, leader selection, velocity and position updates, repair and local search operators, and the stopping condition.

https://doi.org/10.31436/iiumej.v27i2.4072

```
Input:
    - Set of demand points I, candidate facilities J
    - Distance matrix D = [d_ij], fixed costs c_j, demand weights w_i
    - Parameters: P (max facilities), r (service radius), T (max iterations)
    - M (population size), A (archive size), w_init, w_final, c_1, c_2
Output:
    - External archive A* containing non-dominated solutions
Begin
    1. Initialize the population of M particles:
        For each particle:
            - Randomly initialize binary position vector y ∈ {0,1}^|J| s.t. ∑y_j ≤ P
            - Set velocity vector v ← 0
            - Derive x_ij and z_i deterministically from y
            - Evaluate objectives F_1, F_2, F_3
            - Set personal best p_j ← y
    2. Initialize external archive A* with non-dominated solutions from the population
    3. For iteration t = 1 to T do:
        3.1 Update inertia weight:
            w_t ← w_init − (w_init − w_final) × (t / T)
        3.2 For each particle, do:
            a. Select leader g_j from A* using crowding-distance roulette
            b. Update velocity for each dimension j:
                v_j ← w_t × v_j + c_1 × rand1 × (p_j − y_j) + c_2 × rand2 × (g_j − y_j)
            c. Apply sigmoid transformation:
         s_j ← 1 / (1 + exp(−v_j))
            d. Update position y_j:
                y_j ← 1 if rand( ) < s_j else 0
            e. Apply feasibility repair:
                - If ∑y_j > P, remove facilities based on the utility ratio
                - Ensure every demand point is assigned, and the coverage constraint holds
            f. Derive x_ij and z_i from y
            g. Evaluate objectives F_1, F_2, F_3
            h. Update personal best if current solution dominates p_j
        3.3 Update archive A*:
            - Insert non-dominated solutions from the population
            - Remove dominated solutions
            - If |A*| > A, prune using the crowding distance
        3.4 Apply local search to a subset of A*:
            - Use swap, add/drop, or reassignment moves to refine selected solutions
    End For
Return archive A* containing an approximated Pareto front
End
```

Figure 1. MOPSO-FLP pseudo code.

3.3. Specialized Operators

To maintain feasibility with respect to the facility limit and coverage constraints, MOPSO-FLP uses a repair operator whenever a particle violates the upper bound on the number of facilities. Facilities are ranked according to a utility measure that combines their contribution to coverage and their fixed costs, and the least useful facilities are closed until the limit is met.

Local search procedures are applied to selected non-dominated solutions in the archive. Typical moves include swapping facilities (open one, close another), reassigning demand nodes between facilities, and add-drop operations. These moves refine solutions without losing the archive's overall diversity. Diversity is explicitly monitored using crowding distance and related measures, and the archive is pruned when it becomes too large by removing overly similar solutions in objective space.

https://doi.org/10.31436/iiumej.v27i2.4072

3.4. Experimental Design

To evaluate the proposed algorithm, both benchmark and synthetically generated instances are used. Standard instances from the OR-Library provide comparability with prior studies and allow indirect comparison with existing results. Synthetic FLP instances are constructed to vary systematically in size and key characteristics, including the number of demand points and candidate sites, their spatial distribution, demand heterogeneity, and cost structures. The main characteristics of the synthetic instances are summarized in Table 1, which groups them into small, medium, and large categories.

Table 1. Characteristics of generated facility location problem instances

Instance Type	Demand Points (I)	Candidate Locations (J)	Spatial Distribution	Demand Distribution	Cost Structure
Small	50-100	10-20	Uniform	Uniform	Low/High fixed cost
Medium	200-300	30-50	Clustered	Normal	Mixed ratios
Large	400-500	80-100	Random	Exponential	Variable ratios

For each instance, MOPSO-FLP and the comparator algorithms are run multiple times using different random seeds. Performance is summarized using three common multi-objective indicators: hypervolume, inverted generational distance, and spacing. Wall-clock time is also recorded to gauge computational effort.

Comparisons are made against Non-dominated Sorting Genetic Algorithm II (NSGA-II), a multi-objective simulated annealing variant (MO-SA), a greedy multi-objective heuristic (GMOH), and a variable neighborhood search for multi-objective optimization (VNS-MO). All algorithms share similar stopping rules, such as a maximum number of iterations and a limit on consecutive iterations without improvement, to provide a fair basis for comparison.

3.5. Implementation Details

The algorithm is implemented in C++ using a modular architecture that separates problem representation, objective evaluation, and search operators. Objective evaluations are parallelized using multi-threading to speed up runs on multi-core hardware. Experiments are conducted on a workstation with an Intel Core i7-8700K processor, 32 GB of RAM, and Windows 10 Professional, with compiler optimization flags enabled.

Table 2. Experimental parameter settings for MOPSO-FLP

Parameter	Value / Range	Justification
Population size	100 particles	Balances exploration and computational efficiency
Maximum iterations	1000	Provides sufficient convergence for large instances
Archive size	100 solutions	Prevents archive overflow while maintaining diversity
Inertia weight	$0.9 \rightarrow 0.1$	Decreases nonlinearly to promote exploration, then exploitation
Acceleration coefficients	$c_1 = c_2 = 2.0$	Standard PSO setting with adaptive adjustment
Mutation rate	0.01	Introduces diversity and prevents stagnation

Parameter values such as population size, number of iterations, archive size, initial and final inertia weights, acceleration coefficients, and mutation rate are selected based on preliminary experiments and insights from the literature. The chosen settings aim to balance solution quality and runtime (see Table 2).

https://doi.org/10.31436/iiumej.v27i2.4072

3.6. Statistical Analysis Framework

To assess the robustness and statistical significance of the results, each algorithm–instance combination is run 30 times. For each performance indicator, sample means and standard deviations are reported. Non-parametric tests such as the Mann–Whitney U test and Kruskal–Wallis test are used to compare algorithms, and corrections for multiple comparisons are applied when necessary. Effect sizes (e.g., Cohen's d or Cliff's delta) are also calculated to judge the practical relevance of observed differences.

3.7. Validation

Algorithmic correctness is checked through unit tests for core components such as the velocity update, repair operations, and archive management, and through integration tests on small instances where high-quality or optimal solutions are known. Pareto fronts obtained in these tests are inspected for consistency with the model's trade-off structure before running large-scale experiments.

4. Results and Analysis

4.1. Benchmark Performance

MOPSO-FLP was tested on small (50×10), medium (150×30), and large (500×100) instances. Results confirmed consistent performance across scales. The summary statistics for hypervolume, inverted generational distance, spacing, and runtime across the three instance groups are reported in Table 3.

- Small problems: HV = 0.847 (±0.023), IGD = 0.011 (±0.004), SP = 0.042 (±0.010). Pareto fronts were nearly optimal in 87% of runs.
- Medium problems: HV = 0.792 (±0.031), IGD = 0.034 (±0.008), SP = 0.058 (±0.011). Runtimes scaled to ~392s.
- Large problems: HV = 0.718 (±0.042), IGD = 0.067 (±0.015), SP = 0.071 (±0.014). Runtimes ~847s, feasible for planning tasks.

Table 3. Benchmark performance of MOPSO-FLP

Scale	HV (±SD)	IGD (±SD)	SP (±SD)	Time (s)
Small	0.847±0.023	0.011±0.004	0.042±0.010	134±29
Medium	0.792±0.031	0.034±0.008	0.058±0.011	392±71
Large	0.718±0.042	0.067±0.015	0.071±0.014	847±127

To ensure the reliability and statistical significance of the results, each algorithm is executed 30 times per instance.

4.2. Convergence and Trade-offs

Convergence followed a three-phase trajectory: rapid early improvements, steady mid-phase gains, and late-stage stabilization. Final Pareto sets averaged 70–75 solutions, including extreme and knee points.

Trade-off slopes indicated that each 1% service improvement required ~1.34% higher cost, while each 1% equity gain required ~2.17% additional cost. Service and equity were positively correlated ($\rho \approx 0.62$), meaning higher service generally enhanced equity, though with diminishing returns.

https://doi.org/10.31436/iiumej.v27i2.4072

4.3. Comparative Evaluation

MOPSO-FLP was benchmarked against NSGA-II, MOSA, VNS-MO, and GMOH (30 runs each).

- Small problems: MOPSO-FLP (HV = 0.847) outperformed NSGA-II (0.821) and GMOH (0.692).
- Medium problems: MOPSO-FLP (0.792) exceeded NSGA-II (0.751) and VNS-MO (0.718).
- Large problems: MOPSO-FLP (0.718) surpassed VNS-MO (0.673) and NSGA-II (0.651).

Diversity was best preserved (SP = 0.067 vs. 0.089 for NSGA-II), and runtime scaled better than NSGA-II due to reduced dominance-sorting overhead.

4.4. Sensitivity Analysis

Parameter sensitivity indicated that the best performance was achieved with 100 particles, an archive size of 100, and a linearly decreasing inertia weight. Larger populations gave a slight hypervolume (HV) gains but nearly doubled runtime, reflecting diminishing returns and confirming earlier insights on swarm optimization efficiency.

Problem sensitivity indicated that clustered demand improved HV (0.834) compared with uniform (0.792) and random (0.751) distributions. Greater demand heterogeneity also reduced solution quality, with HV dropping from 0.823 to 0.756 as the coefficient of variation rose from 0.2 to 0.8.

4.5. Deployment Validation in a National Postal Logistics Network

The National Post, a state-owned logistics provider, faces rising e-commerce volumes and the challenge of serving both dense urban centers and remote rural areas. To evaluate MOPSO-FLP in a real-world setting, 89 existing post offices were modeled as candidate facilities (m = 89) and 247 districts as demand nodes (n = 247), see Fig. 2. The objectives mirrored strategic priorities: minimize operational costs, maximize service coverage, and reduce maximum delivery times.

4.5.1. Data Preparation

Data transparency and confidentiality: Deployment validation relies on operational planning data from the national postal logistics context. To protect confidentiality, identifiers and sensitive operational attributes were anonymized and, where necessary, aggregated to planning zones; the analysis uses derived distance/time matrices and normalized cost parameters consistent with the operator's planning practice. While the raw dataset cannot be publicly released due to organizational restrictions, all algorithm parameters, evaluation settings, and the facility-location instance structure are fully reported to support replication using comparable datasets or synthetic instances.

Deployment validation was based on an integrated dataset that combined geographic attributes, demand patterns, cost information, and operational constraints. Geographic data were drawn from the National Post's internal GIS and provided GPS coordinates for candidate facility sites and district centers. A road-distance matrix was constructed using the Google Maps API to reflect actual routes and physical barriers, including rivers and mountainous terrain. Demand records for 2022–2023 reported annual parcel volumes at the district level and indicated pronounced heterogeneity, with a coefficient of variation of 0.67, highlighting the contrast between dense urban areas and sparsely populated rural districts. Cost information

https://doi.org/10.31436/iiumej.v27i2.4072

covered both fixed operating expenditures at facilities and variable transportation costs; all values were converted to an annual basis to allow consistent comparison across network configurations. In addition, a maximum delivery time of 5 hours was adopted as a soft service-level constraint, aligned with the National Post's operational guidelines, to explicitly embed realistic delivery speed expectations into the optimization model.

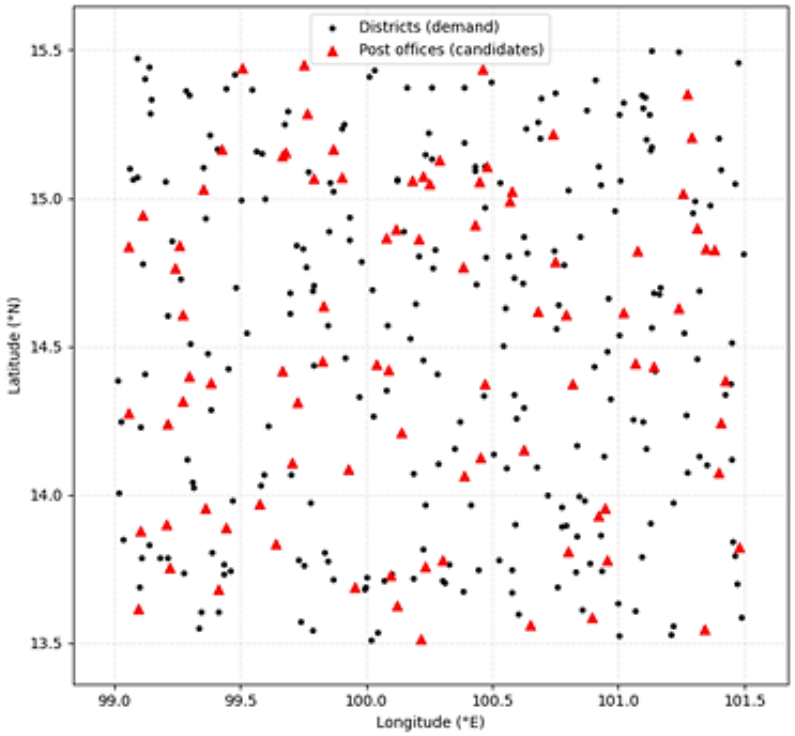


Figure 2. A National Post central region schematic diagram.

4.5.2. Optimal Results

The MOPSO-FLP search produced a Pareto front comprising 47 non-dominated network designs, providing the National Post with a broad portfolio of candidate configurations. Within this set, three representative solutions were identified as particularly illustrative of the key trade-offs between cost, service coverage, and delivery equity.

One extreme of the frontier corresponds to a cost-focused configuration with 23 open facilities and an annual operating cost of US$ 3.54 million. This design achieves 89% service coverage and a maximum delivery time of 4.7 hours. Although financially attractive, it leaves several remote districts relatively weakly served and therefore does not align well with equity-oriented objectives.

Interpretation of spatial patterns and equivalent solutions: For a fixed number of facilities, multiple location configurations may achieve very similar objective values because demand clusters and geographic symmetry can create near-equivalent alternatives. To support implementation, we recommend selecting representative solutions using a transparent tie-break rule, for example: (i) robustness under demand or travel-time perturbations, (ii) reuse of existing sites and ease of staffing, and (iii) regional fairness considerations (e.g., avoiding systematic under-service of remote subregions). In general, cost-focused solutions tend to concentrate facilities near high-demand centers, whereas service-oriented solutions add peripheral sites to reduce maximum travel time and improve coverage; balanced solutions combine a central backbone with a small number of strategically placed edge facilities.

https://doi.org/10.31436/iiumej.v27i2.4072

At the opposite end is a service-oriented configuration that expands the network to 34 facilities. In this case, annual operating costs increase to US$4.36 million, but service coverage rises to 98% and the maximum delivery time falls to 2.1 hours. The configuration is clearly superior from a service perspective, yet the additional cost makes it difficult to justify as a long-term baseline.

Table 4. Key deployment validation solutions

Configuration	Facilities	Cost (M US$)	Coverage (%)	Max Delivery (h)
Cost-efficient	23	3.54	89.2	4.7
Service-optimized	34	4.36	97.8	2.1
Balanced	28	3.83	94.3	3.2

Between these two extremes lies a balanced configuration with 28 facilities that offers a more even compromise across the three objectives. It incurs an annual cost of US$ 3.83 million, delivers 94% service coverage, and limits the maximum delivery time to 3.2 hours. This design was selected by the National Post for pilot implementation. Monitoring after deployment indicated a 12.3% reduction in average delivery times and an 8.7% increase in service coverage, achieved with only a 4.2% rise in operating costs. Customer satisfaction also improved, increasing from 7.2 to 8.1 on a 10-point scale (see Table 4).

4.5.3. Managerial Insights

The managerial insights derived from the National Post deployment highlight several important patterns. First, optimized networks achieved facility utilization rates between 70% and 85%, which balanced efficiency with resilience by preventing both underutilization and capacity saturation. Second, the results confirmed a clear pattern of diminishing returns, as extending service coverage beyond 95% required disproportionately higher operational costs. Third, geographic constraints, such as rivers and mountain ranges, strongly shaped the configuration of service clusters, underscoring the value of incorporating geospatial information into facility location models. Finally, the scalability of MOPSO-FLP was confirmed, as the algorithm solved a national-scale problem in less than 30 minutes, demonstrating its feasibility for integration into real operational planning cycles.

4.5.4. Deployment Summary

The National Post deployment validated that MOPSO-FLP not only generates theoretically optimal solutions but also delivers actionable strategies. The improvements in delivery efficiency, service coverage, and customer satisfaction demonstrate its practical value for national-scale logistics.

4.6. Robustness and Component Analysis

The ablation experiments indicate that adaptive parameter control, the local search phase, and the use of a crowding-distance-based archive all play a substantial role in MOPSO-FLP's performance, together yielding an improvement in hypervolume of almost 20%. In terms of robustness, the algorithm continues to perform satisfactorily under 20% perturbations in distances and 30% variability in demand, with hypervolume values remaining above 0.74. Scalability tests further show that the approach can handle instances with up to 1,000 demand points and 200 candidate facilities; although runtimes exceed 2,300 seconds at this upper end, they remain acceptable for networks with 500–1,000 nodes, which is typical of national-scale logistics systems.

4.7. Statistical Validation

Across 30 runs, HV was bounded within [0.771, 0.813], IGD within [0.029, 0.039], and SP within [0.061, 0.073]. ANOVA ($F = 23.7, p < 0.001, \eta^2 = 0.34$) confirmed significant performance differences. Tukey's HSD identified MOPSO-FLP as superior to all comparators. Non-parametric Kruskal–Wallis and Mann–Whitney U tests supported the same conclusion, with large effect sizes (Cohen's d > 0.8). These results establish that MOPSO-FLP's improvements are statistically significant and practically meaningful.

In summary, MOPSO-FLP demonstrated consistent superiority in benchmark tests, robust trade-off exploration, and statistical reliability. Its resilience to uncertainty and ability to solve national-scale problems, as demonstrated in a postal logistics deployment, confirm its scalability and practical relevance. Future work will explore hybridization with exact methods and dynamic extensions incorporating real-time data to further enhance adaptability for complex logistics and infrastructure systems.

5. Conclusion and Future Work

This work presented a discrete multi-objective particle swarm optimization framework (MOPSO-FLP) tailored to facility location problems that must balance cost efficiency, service coverage, and spatial equity. The method combines a discrete representation of facility decisions with adaptive parameter control and targeted feasibility-repair operators, enabling an effective search of complex Pareto fronts. In benchmark tests, the proposed algorithm consistently outperformed reference methods, including NSGA-II and VNS-MO. Statistical analysis further indicated that MOPSO-FLP is both robust and scalable, supporting its use in large-scale facility location applications.

The application of the framework to a national postal logistics network confirmed its practical value, revealing a rich set of trade-off solutions and underscoring the advantages of balanced designs that enhance delivery performance and equity while keeping additional costs moderate. The same modeling principles can be transferred to other settings, including healthcare provision, smart-city services, and environmentally oriented infrastructure planning. Promising directions for further work include combining the metaheuristic with exact optimization components, incorporating real-time operational data, and embedding equity considerations within participatory planning processes so that MOPSO-FLP remains a versatile tool for multi-objective infrastructure optimization.

Acknowledgement

This research project was granted by the Silpakorn University Research, Innovation, and Creative Fund, 2025. We express our gratitude to the anonymous reviewers for their meticulous examination of our work and their numerous astute observations and recommendations.

References

[1] Amin-Tahmasbi H, Sadafi S, Ekren BY, Kumar V. (2023) A multi-objective integrated optimisation model for facility location and order allocation problem in a two-level supply chain network. Annals of Operations Research, 324:993–1022. https://doi.org/10.1007/s10479-022-04635-1

[2] Goodarzian F, Hosseini-Nasab H, Fakhrzad MB. (2020) A multi-objective sustainable medicine supply chain network design using a novel hybrid multi-objective metaheuristic algorithm. International Journal of Engineering, 33(10):1986–1995.

[3] Mirghaderi SD, Modiri M. (2021) Application of meta-heuristic algorithm for multi-objective optimization of sustainable supply chain uncertainty. Sādhanā, 46:52. https://doi.org/10.1007/s12046-020-01554-4

[4] Yan T, Lu F, Wang S, Wang L, Bi H. (2023) A hybrid metaheuristic algorithm for the multi-objective location-routing problem in the early post-disaster stage. Journal of Industrial and Management Optimization, 19(6):1. https://www.aimsciences.org/article/doi/10.3934/jimo.2023104

[5] Pourghader CA, Sabk Ara M, Moradi Pirbalouti S, Khadem M, Bahrami S. (2022) A multi-objective location-routing problem model for multi-device relief logistics under uncertainty using meta-heuristic algorithm. Journal of Applied Research on Industrial Engineering, 9(3):354–373. https://doi.org/10.22105/jarie.2021.299798.1365

[6] Hemici M. (2024) Multi-objective optimization for supply chain management. PhD thesis, Université de Bordj Bou Arreridj, Faculty of Mathematics and Computer Science.

[7] Suryawan RF, Fatchoelqorib M, Septiano R, Sari L, Widodo S, Yosepha SY, Devi NK. (2022) Two meta-heuristic algorithms for solving multi-objective model for the service quality and price in the digital supply chain. Industrial Engineering and Management Systems, 21(3):440–448. https://doi.org/10.7232/iems.2022.21.3.440

[8] Fugaro S, Sgalambro A. (2025) Multi-objective facility location problems: Systematic literature review and research agenda. Available: https://doi.org/10.2139/ssrn.5178615 (accessed 3 June 2025). https://ssrn.com/abstract=5178615

[9] Arevalo-Ascanio R, De Meyer A, Gevaers R, Guisson R, Dewulf W. (2024) Location-routing problem for integrated supply chain network design with first and last mile: A critical literature review. Operations and Supply Chain Management: An International Journal, 17(2):206–219. https://doi.org/10.31387/OSCM0570423

[10] Koosha H, Mirsaeedi F, Assadi MT. (2025) A multi-objective genetic algorithm for unequal area facility layout problem considering safety and cost. Soft Computing, 29:3869–3887. https://doi.org/10.1007/s00500-025-10605-z

[11] Liu M, Song Q, Zhao Q, Li L, Yang Z, Zhang Y. (2022) A hybrid BSO–ACO for dynamic vehicle routing problem on real-world road networks. IEEE Access, 10:118302–118312. https://doi.org/10.1109/ACCESS.2022.3221191

[12] Cakmak E, Önden İ, Acar AZ, Eldemir F. (2021) Analyzing the location of city logistics centers in Istanbul by integrating geographic information systems with binary particle swarm optimization algorithm. Case Studies on Transport Policy, 9(1):59–67. https://doi.org/10.1016/j.cstp.2020.07.004

[13] Makhadmeh SN, Kassaymeh S, Rjoub G, Bataineh B, Sanjalawe Y, Al-Betar MA. (2025) Recent advances in multi-objective whale optimization algorithm, its versions and applications. Journal of King Saud University – Computer and Information Sciences, 37(7):1–38. https://doi.org/10.1007/s44443-025-00184-2

[14] Kaveh A, Javid AAS, Vazirinia Y. (2025) Multi-objective variants of water strider algorithm for construction engineering optimization problems. Periodica Polytechnica Civil Engineering, 69(3):925–943. https://doi.org/10.3311/PPci.40442

[15] Zhang W, Xiao G, Gen M, Geng H, Wang X, Deng M, Zhang G. (2024) Enhancing multi-objective evolutionary algorithms with machine learning for scheduling problems: Recent advances and survey. Frontiers in Industrial Engineering, 2:1337174. https://doi.org/10.3389/fieng.2024.1337174

[16] Ezugwu AES. (2023) Metaheuristic optimization for sustainable unrelated parallel machine scheduling: A concise overview with a proof-of-concept study. IEEE Access, 12:3386–3416. https://doi.org/10.1109/ACCESS.2023.3347047

[17] Alqahtani H, Kumar G. (2024) Efficient routing strategies for electric and flying vehicles: A comprehensive hybrid metaheuristic review. IEEE Transactions on Intelligent Vehicles, 9(9):5813–5852. https://doi.org/10.1109/TIV.2024.3358872

https://doi.org/10.31436/iiumej.v27i2.3968

Optimum Alert For In-Vehicle Safety Technologies

NOOR IRZA MOHD ZAKI[1*], MOHD KHAIRI ABU HUSAIN[1], NURULAKMAR ABU HUSAIN[2], SITI MAHIRAH CHE HUSIN[3], YAHAYA AHMAD[4]

[1]Faculty of Artificial Intelligence, Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia
[2]Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia
[3]ACTS Smart Solutions Sdn Bhd, Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia
[4]Malaysian Institute of Road Safety Research (MIROS), 43000 Kajang, Selangor, Malaysia

**Corresponding author: noorirza.kl@utm.my*

(Received: 2 October 2025; Accepted: 26 January 2026; Published online: 10 May 2026)

ABSTRACT: Vehicle safety technologies have played a critical role in protecting vehicle drivers for several decades. Auditory signals, for example, convey relevant data to the driver without interfering with the driving task. Vehicle auditory alerts help ensure the safest driving experience possible by providing warnings, thereby reducing road accidents caused by reckless or careless driving. Nonetheless, auditory alerts have long been a source of contention in vehicle safety features due to driver annoyance and false-alarm issues. This study aims to investigate drivers' awareness and perceptions of in-vehicle safety technology and auditory alerts, and how these factors affect drivers' reactions to the driving environment. A structured questionnaire was designed and distributed to 150 licensed drivers to collect data on their exposure to safety features, perceived usefulness and urgency of alerts, and preferred auditory characteristics. Descriptive and comparative analyses were conducted to evaluate drivers' perceptions alongside benchmark data from Original Equipment Manufacturers (OEMs). Results show that 87% of respondents are aware of major in-vehicle safety technologies, yet 48.7% perceive them as potentially distracting while driving. Findings also identify 1000 Hz at 65 dB as the preferred auditory frequencies for safety alerts based on user responses. This study provides insights for policymakers, manufacturers, and safety stakeholders to enhance the design, acceptance, and effectiveness of in-vehicle auditory warning systems.

ABSTRAK: Teknologi keselamatan kenderaan memainkan peranan penting dalam melindungi pemandu melalui ciri amaran seperti isyarat auditori yang menyampaikan maklumat keselamatan tanpa mengganggu tugas pemanduan. Walaupun amaran auditori dapat membantu mengurangkan risiko kemalangan jalan raya, isu seperti gangguan kepada pemandu dan amaran palsu masih menjadi cabaran utama. Kajian ini meneliti tahap kesedaran dan persepsi pemandu terhadap teknologi keselamatan dalam kenderaan serta amaran auditori, termasuk kesannya terhadap tindak balas pemandu dalam persekitaran pemanduan. Data diperoleh melalui soal selidik berstruktur yang melibatkan 150 pemandu berlesen, merangkumi pendedahan terhadap ciri keselamatan, persepsi kegunaan, tahap kecemasan amaran dan pilihan ciri auditori. Analisis deskriptif dan perbandingan turut dijalankan berdasarkan data penanda aras daripada Original Equipment Manufacturers (OEM). Dapatan menunjukkan bahawa 87% responden menyedari kewujudan teknologi keselamatan utama dalam kenderaan, namun 48.7% menganggapnya berpotensi mengganggu pemanduan. Kajian ini juga mengenal pasti bahawa frekuensi 1000 Hz pada 65 dB merupakan ciri auditori yang paling digemari untuk amaran keselamatan. Dapatan ini dapat membantu pembuat dasar,

pengeluar kenderaan dan pihak berkepentingan meningkatkan reka bentuk, penerimaan dan keberkesanan sistem amaran auditori dalam kenderaan.

KEYWORDS: *Vehicle Safety, In-Vehicle Technology, Auditory Alert, Awareness and Perception, Optimum Sound Characteristics.*

1. INTRODUCTION

Advances in technology have accelerated in recent years, with the automotive sector undergoing significant transformations in line with the emergence of the Fourth Industrial Revolution (IR 4.0). This transformation encompasses various aspects of industrial processes, including design, manufacturing, management, and vehicle maintenance. Automakers and Original Equipment Manufacturers (OEMs) are rapidly adapting their business models to align with Industry 4.0 principles to remain relevant and competitive in the global automotive industry. At the same time, they are also trying to convince customers to adopt innovative car technology. Vehicle safety technologies, for example, are used to alert and warn drivers of potentially hazardous circumstances.

According to the World Health Organization (WHO) in its Global Status Report on Road Safety 2018 [12], the number of road accident fatalities has increased gradually over the years, reaching 1.35 million in 2016. With the significant number of yearly fatalities and injuries caused by traffic accidents, the United Nations (UN) emphasized two goals in the 2030 Agenda for Sustainable Development: 1) a 50% decrease in worldwide road traffic fatalities and injuries by 2030; and 2) a safe, economical, and sustainable transportation system for all by 2030 [13]. These goals align with the WHO's Global Plan for the Decade of Action for Road Safety 2011-2020 [14].

In the Decade of Action for Road Safety [14], WHO suggested five road safety pillars: road safety management, safer mobility and roads, safer cars, safer road users, and post-crash management. All countries must enhance road safety holistically and implement effective interventions. Human error, vehicle and equipment failures, and road environmental conditions were identified as the leading causes of road fatalities in studies by Hamid et al. and Mohammed et al. [16, 17].

Advances in technology and information about vehicle safety systems have opened new opportunities to reduce traffic accidents and increase drivers' awareness of their surroundings. As a result, automakers also put more effort into creating new vehicles with advanced safety features [22]. However, each vehicle manufacturer refers to the technology by a different name, even though it operates in the same manner. Vehicle safety technology may communicate with the driver through various modalities, including visual, auditory, and haptic alerts. The combination of visual and auditory alerts is more effective, providing the driver with better situational awareness than the visual modality alone [23-26].

Consequently, automakers and OEMs must adhere to crashworthiness safety performance evaluation criteria and operational safety technology requirements. Furthermore, according to the ASEAN NCAP Labelling Guideline for Manufacturers (applicable only in Malaysia), all automakers and OEMs must incorporate vehicle safety technology, i.e., airbags, Anti-Lock Braking System (ABS), Electronic Stability Control (ESC), Blind-Spot Detection, Lane Keep Assist, Lane Departure Warning, Forward Collision Warning, Seat Belt Reminder (SBR) and Autonomous Emergency Braking (AEB) [18].

https://doi.org/10.31436/iiumej.v27i2.3968

Driver-assistance systems that warn of potential collisions have become standard features in modern vehicles. Today's advanced automotive technologies rely on a network of sensors and cameras to identify nearby vehicles and other obstacles. When the likelihood of a collision surpasses a defined safety threshold, the system issues a warning to alert the driver. Research has shown that the effectiveness of these alerts can be undermined by frequent false alarms and mismatches between the perceived urgency of a warning and the actual severity of the situation [1, 2].

Perceived urgency strongly influences how quickly a driver responds, but increasing urgency often makes alerts more annoying, creating a well-recognized trade-off between urgency and annoyance [3, 4]. The context in which the signal is perceived, on the other hand, affects this relationship [5]. According to some studies, the characteristics of auditory alerts result in varying degrees of urgency [6, 7]. The emotional state of the driver (i.e., alert, fatigue, etc.), the driver's age, driving experience, driving conditions (i.e., driving during busy traffic hours or at nighttime with minimal traffic), physical design, innovation, and auditory features (i.e., design of the alarms) all contribute to the irritation of the warning systems [8, 9].

A good urgency alert helps drivers understand warnings and respond appropriately, as they are more likely to respond to high-urgency alerts. Despite this, high-urgency alerts tend to reduce response efficacy due to increased false alarms [10, 11]. Previous studies consistently show that poorly designed auditory alerts can irritate drivers, leading some to ignore or turn off safety warnings altogether. When the pulse repetition rate, the fundamental frequency, and several repeats operate in unison, this substantially affects driving behavior [6].

Edworthy & Hellier [28] identified four attributes of the ideal auditory alert: easily localizable in space, not susceptible to masking by other sounds, not a source of interference for other communications (e.g., speech), and user-friendly and easily manageable. Nevertheless, the auditory alert has long been a contentious topic in vehicle safety technology, primarily due to driver annoyance and false-alarm issues. A survey conducted by Block et al. [29] reported that 70% of drivers turned off the warning alarms due to annoyance with the audio characteristics. Sayed et al. [30] revealed that drivers often feel startled and annoyed when warning alarms are triggered during non-emergency situations that do not require immediate intervention. Similarly, a J.D. Power consumer survey [31] found that many drivers found such warning alerts to be bothersome, leading some to turn off the alert system entirely. This tendency raises serious concerns regarding the effective implementation of vehicle safety technologies.

Malaysia is an essential context for this research because it has one of the highest motorization rates in Southeast Asia. Hence, the Malaysian Institute of Road Safety Research (MIROS), in collaboration with the New Car Assessment Program for Southeast Asian Countries (ASEAN NCAP), has implemented several programs, including developing related guidelines/protocols to improve road safety, raise vehicle safety standards, and promote safer vehicles. For example, the Malaysian Ministry of Transport is constructing a framework for the Malaysian Road Safety Plan 2021-2030 [15].

Thus, this study aims to investigate drivers' awareness and perceptions of vehicle safety technologies with auditory alerts. Towards the end of this study, the awareness and perceptions of Malaysian drivers regarding existing vehicle safety technologies with auditory alerts are discussed, and the preferred frequency level of auditory alerts is mapped to align with the selected OEMs' alert frequency settings.

In this study, the technologies were chosen based on the auditory-alert modality in Advanced Driver Assistance Systems (ADAS) and other Intelligent Transport Systems (ITS).

Based on the European Automobile Manufacturers' Association (ACEA) Code of Practice for ADAS development, drivers have two primary tasks: maneuvering or guiding the vehicle, and stabilizing and navigating. Hence, the ADAS features are designed to support the driver in their primary task and to assist with at least either stabilization or navigation. [27].

The National Highway Traffic Safety Administration (NHTSA) categorizes driver assistance technologies into four main groups: (i) braking and collision avoidance, (ii) backup and parking assistance, (iii) lane and side assistance, and (iv) systems for maintaining a safe following distance [32]. Moravčík and Jaśkiewicz [33] further distinguish intelligent vehicle safety technologies according to their operational timing relative to a crash event: (i) systems that continuously support the driver's actions, (ii) systems that activate at the moment of impact, and (iii) systems that function after a collision. To facilitate understanding, Table 1 organizes 12 vehicle safety technologies that incorporate auditory alerts into five distinct categories.

Table 1. Classification of vehicle safety technology (with auditory alert).

Category	**Vehicle Safety Technology (with Auditory Alert)**
Collision Warning	Pre-Collision Warning (Forward & Reverse), Pedestrian Detection Warning (PDW), Lane Departure Warning (LDW), Blind Spot Monitoring (BSM), Rear Cross Traffic Alert (RCTA)
Collision Intervention (Braking)	Collision Mitigation Brake System (CMBS)
Parking Assistance	Parking Assistance Technology
Driving Control Assistance	Adaptive Cruise Control (ACC), Lane Keeping Assistance System (LKAS)
Other Safety Reminder Technology	Seat Belt Reminder (SBR), Door Ajar/Door Left Open Warning System, Tire Pressure Warning System

In this study, the technologies were organized by their functional support to drivers: collision warning, collision intervention, driving control assistance, parking assistance, and safety reminder systems [32, 34-35]. Collision warning features detect potential hazards around the vehicle, while collision intervention systems apply braking or acceleration to prevent or mitigate impact. Driving control assistance helps maintain safe distances and lane position; parking assistance guides low-speed maneuvers, and safety reminder technologies provide persistent alerts until corrective action is taken.

2. METHODOLOGY

The methodology of this study comprises three phases. Phase 1 focuses on developing a comprehensive database of chime technologies [36, 37]. Phase 2 examines drivers' awareness and perceptions of vehicle safety technologies that employ auditory alerts. Phase 3 aims to identify the optimal sound characteristics that promote effective driver responses in real-world driving environments. An overview of these three phases is shown in Figure 1.

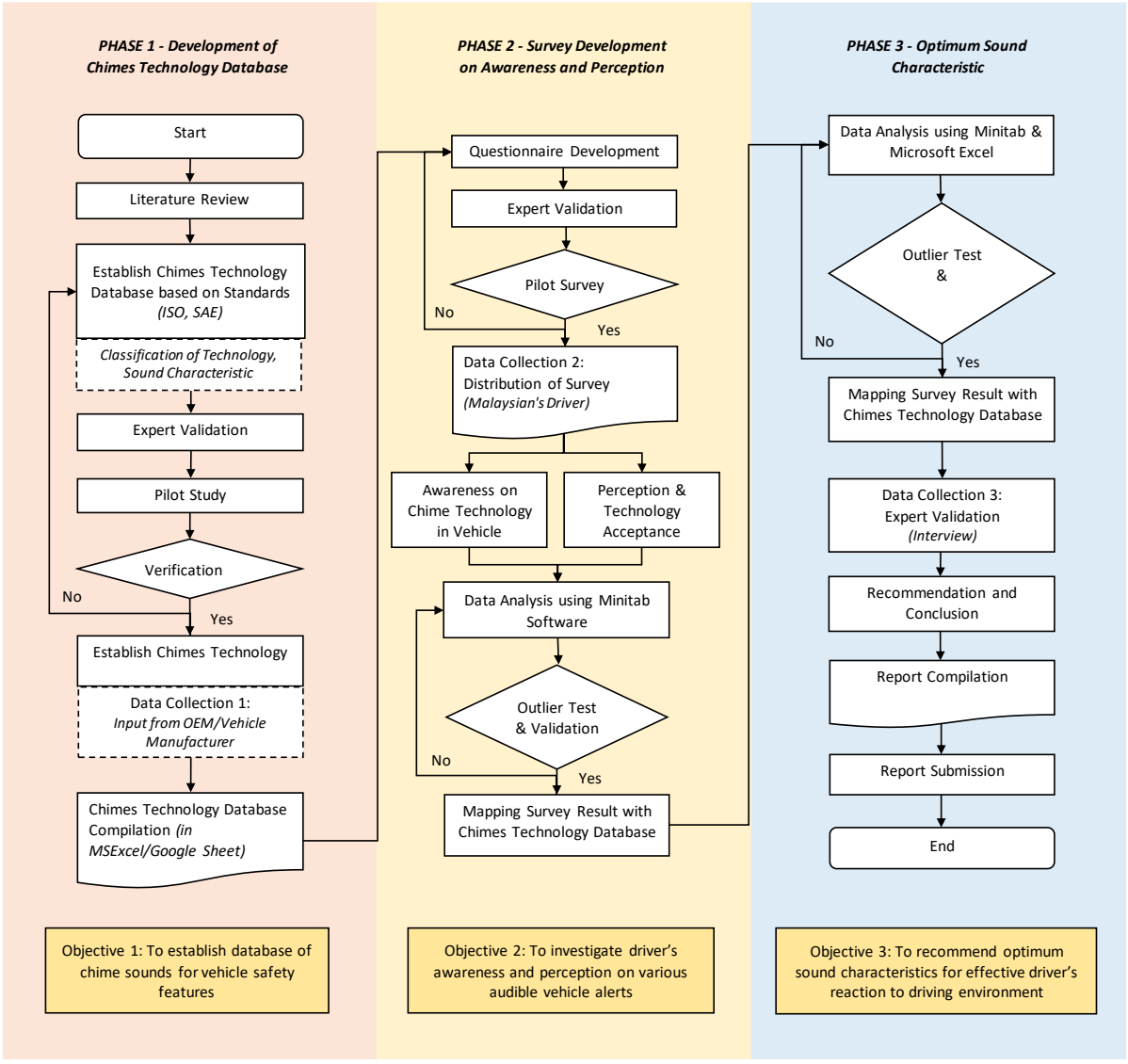


Figure 1. Research methodology flowchart.

2.1. Phase 1: Development of the Chime Technology Database

The chime technology database for this study was developed using proprietary information provided directly by selected OEMs, thus offering authentic insight into the in-vehicle alert systems used in the market. Each OEM was provided with a structured template to specify the details of its in-vehicle auditory alerts, including audio files, dominant frequency characteristics, sound pressure levels, and associated safety functions. In cases where OEMs were unable to share actual alert files due to confidentiality constraints, the researchers generated supplementary auditory samples based on the dominant frequencies recommended in ISO 15006, ensuring consistency across datasets. All collected and generated information was then compiled into a standardized database, which acted as the baseline reference for subsequent analyses of driver preferences.

The development of the chime technology database began with a comprehensive literature review of existing vehicle safety systems that utilize auditory alerts. This review took into account the sound characteristics recommended by the International Organization for Standardization (ISO) [38] and the standards established by the Society of Automotive Engineers (SAE) for each respective technology [34, 35]. Subsequently, a market survey was

conducted using data from the Malaysian Automotive Association [39–45] to provide an overview of Malaysia's top five vehicle brands and their total sales.

The survey revealed that Perodua led the Malaysian market in vehicle sales, followed by Proton, Honda, Toyota, and Nissan [39–45]. This result was used to identify potential OEMs or vehicle manufacturers for data collection on chime technology. Based on this OEM's selection, a standardized database template was distributed to the selected OEMs, requesting detailed information on the vehicle safety technologies in their models that incorporate audible alerts. Each OEM was asked to specify the type of warning modality, provide corresponding audio files, and submit relevant user manuals. The collected inputs were consolidated into a comprehensive database of Chime technology.

2.2. Phase 2: Survey Development on Awareness and Perception

In Phase 2 of this study, a survey approach was employed to assess Malaysian drivers' awareness, perceptions, and acceptance of in-vehicle chime technologies. Before data collection, the survey was validated by experts from the Malaysian Institute of Road Safety Research (MIROS) with at least 10 years of experience to ensure it covered critical aspects and identified any missing components.

The survey was developed using Google Forms and is divided into four parts. Part A: Demographics covers demographic characteristics such as age, nationality, driving experience, car ownership, and type of vehicle driven, including its model year. Part B: Driver's Awareness of In-Vehicle Safety Technology with Auditory Alert focuses on awareness of the chime technology, in which the respondents need to answer the vehicle buying factor, awareness of vehicle safety technology available in their car, drivers' interaction with the technology, and the importance of vehicle safety technologies with auditory alert. In Part C: Driver's Perception of In-Vehicle Safety Technology with Auditory Alert, respondents must rate their perceptions of vehicle safety technology in terms of ease of use, usefulness, and the auditory alert's safety and effectiveness using the Likert scale. Lastly, in Part D: Audio Test, respondents need to select their preferences for perceiving urgency and annoyance for each audio alert.

The auditory alerts used in the survey were configured with distinct characteristics, including frequency, intermittent cycles, and sound pressure levels. They were created using NCH's Tone Generator software. Meanwhile, WavePad Audio Editing software was used to compile the audio files. The audio frequencies used in this study comprised three levels: 500 Hz, 1000 Hz, and 1500 Hz, with corresponding sound pressure levels (SPLs) set at 65 dB, 70 dB, 75 dB, and 80 dB. Additionally, the intermittent sound cycles used in this study were set at 500 ms, 1000 ms, and 2000 ms.

Prior to the main data collection, a pilot test was conducted to ensure that the auditory alert characteristics met the desired specifications and to estimate the time required for respondents to complete the survey. To verify the accuracy of the desired sound pressure levels, a calibrated sound level meter was used.

A probability sampling method was adopted in this study to obtain a representative sample of the population. This is to ensure equal selection probabilities and thereby reduce sampling bias [46]. Based on the sample size calculation method outlined in [47], at least 120 respondents were required. The survey was designed with a 10% margin of error, a 90% confidence level, and a 50% distribution across each category. Only Malaysian drivers with driving experience are allowed to participate as respondents. They were categorized into four age groups: 18 to 24 years old, 25 to 39 years old, 40 to 54 years old, and 55 years old and above [48].

2.3. Phase 3: Optimum Sound Characteristic

This phase aims to identify and map the preferred frequency levels for auditory alerts and align them with the frequency settings provided by the selected OEMs. In this study, the sound characteristics identified from the driver survey were compared with the OEM-provided auditory frequency settings. Dominant frequencies were extracted from the OEM audio samples using Python software, following the frequency guidelines specified in ISO 15006 [38]. The extracted frequencies were then mapped to the warning-priority levels corresponding to different crash-event stages, based on the established classifications [49]. The alert priority was determined according to the sequence of crash events, ranging from everyday driving to emerging or critical situations [49]. Four priority levels were defined, as shown in Table 2, to reflect increasing urgency: Notification, Low, Medium, and High. The chime technology database was subsequently utilized to determine the most suitable frequency range for auditory alerts; hence, it was the basis for recommending optimal sound characteristics that promote effective driver response.

Table 2. Specification of audio warning priority.

Audio Warning Priority (Hz)	Notification	Medium
	Low	501-800
	Medium	501-800
	Medium	501-800

The data collected from Malaysian drivers were processed and analyzed using Microsoft Excel and Minitab Version 20. Excel was used to sort the raw survey data, filter out outliers, and identify the key characteristics of the audio test (frequency, sound pressure level, and intermittent cycle). Minitab software was used to conduct statistical analyses, encompassing descriptive statistics such as means and standard deviations [50]. Data collection sessions were conducted at several locations across the Klang Valley using smartphones, sound level meters, and Bluetooth-enabled vehicles. During each audio test, the sound samples were played from a smartphone connected to the car's audio system via Bluetooth.

3. RESULT AND DISCUSSION

3.1. Phase 1: Development of the Chime Technology Database

Significant differences in the availability of chime-enabled safety technologies across vehicle classes and specification levels were found in the initial desktop study. In general, higher-specification vehicle models were equipped with multiple auditory-based systems, such as collision warning, collision intervention, driver assistance, and safety reminder technologies. On the other hand, limited auditory alert systems were offered in lower-specification models, with some relying solely on the dashboard's visual indicators. This variation highlights an implementation gap that may affect drivers' exposure to important auditory warnings, given that road-crash risk is equally distributed across all vehicle types.

The desktop study also discovered that most OEMs operate within the recommended dominant frequency range of 500–1500 Hz, although several high-frequency alerts exceeded 3000 Hz. Such high frequencies may contribute to driver annoyance, consistent with previous human-machine interaction (HMI) literature. Additionally, OEMs were found to differ substantially in their warning-priority tuning, with some applying high-urgency tones even for non-critical alerts. This observation aligns with the findings of Sayed et al. [30], who reported that drivers are often startled and annoyed by the warning alarm triggered in non-emergency situations. These findings highlight the pertinence of aligning auditory alert design with

urgency–annoyance principles to minimize distraction and enhance user acceptance of in-vehicle safety technologies.

Auditory alerts in in-vehicle safety systems serve an essential communication role in the human–machine interface, guiding and conveying information to drivers while effectively capturing their attention during maneuvering vehicles. As indicated by the Guideline for Safety of In-Vehicle Information Systems and the Human Factors Design Guidance for Driver–Vehicle Interfaces [51, 52], the recommended sound pressure levels are 50–90 dB(A), with an ideal minimum of 75 dB(A) and frequencies between 500–2500 Hz, including at least two dominant frequencies within 500–1500 Hz. This range aligns with the most sensitive region of the human hearing range. Sounds above 90 dB(A) or above 3000 Hz should be avoided because they may startle or irritate drivers. Nevertheless, alerts should also be at least 15 dB(A) higher than ambient noise to remain detectable. Alert urgency corresponds to the stage of a potential collision and is typically grouped into three levels: low (response needed within 10–120 s), medium (2–10 s), and high (0–2 s) [49]. In general, research consistently shows that alerts with higher frequencies, greater intensity, and shorter intervals are perceived as more urgent [49, 51, 53].

Table 3 presents the database of chime-related technologies collected from five OEMs. The database indicates that all OEMs provide Seat Belt Reminder (SBR) and pre-collision Warning systems, while low-specification models from several brands rely solely on visual indicators. Of the 17 safety technologies examined, only Child Presence Detection (CPD) was absent across all OEMs. For other features such as the Master Warning System, Pedal Misoperation Control, and Tire Pressure Warning System, most OEMs provided only visual indicators without accompanying chime alerts.

Several OEMs also employed the same auditory alerts across multiple safety functions. For example, the Collision Mitigation Braking System (CMBS) in OEM-03 and OEM-04 used the same alert tone for several collision-related features, including Reverse Collision Warning, Forward/Pre-Collision Warning, Pedestrian Detection Warning, and Pre-Collision Braking. Only a few technologies, such as the Master Warning System in OEM-02 and the Tire Pressure Warning System in OEM-04, integrated both visual and auditory cues, while most relied exclusively on dashboard icons.

Dominant frequencies from the available chime samples were extracted using Python, following the recommendations in ISO 15006 [38]. These frequencies were subsequently categorized into several warning-priority levels, as presented in Table 2. The analysis revealed significant variation in frequency settings across OEMs, likely influenced by factors such as the vehicle model, recording conditions, and background noise in the submitted audio files.

Overall, OEM-04 predominantly used high-priority frequencies (1600–2400 Hz) across most technologies while remaining within the limits prescribed by ISO 15006. In contrast, OEM-03 used lower-priority tones around 800 Hz for many alerts, except for Adaptive Cruise Control (ACC), Seat Belt Reminder (SBR), and Pedal Misoperation Control. Notably, the Pedal Misoperation Control alert used a 3200 Hz tone, which exceeds the recommended comfort range for human hearing. This finding is consistent with the design recommendations reported by Campbell et al. [51] and Stevens et al. [52], who emphasized that high-frequency tones can contribute to driver discomfort and annoyance. In summary, while the OEM frequency settings generally align with recommended practices, the observed variations highlight the need for further research on Malaysian drivers' acceptance of these auditory alerts.

https://doi.org/10.31436/iiumej.v27i2.3968

Table 3. Database of chimes-related technology.

			Vehicle Brand				
			OEM-01	**OEM-02**	**OEM-03**	**OEM-04**	**OEM-05**
Chime Technologies	**Collision Warning**	Reverse Collision Warning (RCW)	Visual	Visual	800	2000	2400
		Forward/Pre-collision Warning	1000	1300	800	2000	2000
		Pedestrian Detection Warning (PDW)	NA	800	800	2000	Visual & Audio
		Lane Departure Warning (LDW)	650	Visual & Haptic	NA	1600	1000
		Lane Change Decision Aid Systems (LCDAS)/ Blind Spot Monitoring (BSM)	650	Visual	NA	Visual	Visual & Audio
		Rear Cross Traffic Alert (RCTA)	NA	Visual	NA	2400	1000 / 3000
	Collision Intervention	Pre-collision Braking	NA	NA	800	2000	2400
		Autonomous Emergency Braking (AEB)/ Collision Mitigation Braking System (CMBS)	Visual	1300	800	2000	NA
	Driver Control Assistance	Lane Keeping Assistance System (LKAS)/ Lane Tracing Assist (LTA)	Visual	1300	NA	1600	Visual
		Adaptive Cruise Control (ACC)/ Front Departure Alert (FDA)	Visual	Visual	1600	800	Visual & Audio
	Parking Assistance	Parking Assistance System	1500	Visual	Visual	Visual	Visual
	Other Safety Reminder Technology	Seat Belt Reminder	650 / 2400	500 / 800	800 / 3200	600 / 800	Visual & Audio
		Door Ajar/Door Left Open Alarm	650	Visual	800	800	Visual
		Child Presence Detection (CPD)	NA	NA	NA	NA	NA
		Master Warning System	Visual	500 / 1000	Visual	Visual	NA
		Pedal Misoperation Control	NA	NA	800 / 3200	2000	Visual
		Tire Pressure Warning System	Visual	Visual	Visual	800	Visual

Legend:		
Audio Warning Priority (Hz)	Notification	300-500
	Low	501-800
	Medium	801-1200
	High	1201-3000

https://doi.org/10.31436/iiumej.v27i2.3968

3.2. Phase 2: Survey Development on Awareness and Perception

3.2.1. Demographic Data

A total of 150 respondents took part in this survey. In the 18 to 55 age group, the highest percentage is 45.0%, comprising drivers aged 25-39 years with more than 10 years of driving experience. Participants were mostly from the Klang Valley area (93.3%), while the remaining 6.7% resided outside the Klang Valley. Such a result is expected, as the Klang Valley has a higher population density and a higher concentration of vehicle ownership.

The demographic result also indicates that 71.3% of respondents own a car, with 65.1% owning a national brand. On the other hand, 28.7% of the respondents who do not own a car were drivers with less than 5 years of driving experience (36 respondents), followed by 6-10 years (5 respondents) and more than 10 years (2 respondents). In addition, 39.8% of the respondents owned the latest vehicle model (2016-present). Many respondents also owned either a sedan or a compact car (42.6%). The high percentage of national car users reflects Malaysia's current market distribution, thereby reinforcing the relevance of analyzing locally standard vehicle models. Table 4 tabulates the demographic characteristics of the survey participants.

Table 4. Demographic data of the survey participants.

Variables		Percentage (%)
Age:	18 – 24 years	33.3
	25 – 39 years	44.7
	40 – 54 years	16.8
	≥ 55 years	5.3
Gender:	Male	52.3
	Female	47.7
Current location:	Klang Valley	93.3
	Outside Klang Valley	6.7
Driving experience :	1 – 5 years	39.3
	6 – 10 years	16.0
	>10 years	44.7
Car's possession:	Yes	71.3
	No	28.7
Car's model	National car	34.9
	Imported car	65.1
Car's model year:	Before 2005	9.3
	2006 – 2010	21.3
	2011 – 2015	29.6
	2016 – Present	39.8
Car's type:	Sedan	42.6
	Compact	42.6
	Multi-Purpose Vehicle (MPV)	10.2
	Sport Utility Vehicle (SUV)	4.6

3.2.2. Driver's Awareness of Various Audible Alerts

Figure 2 illustrates the key factors drivers will consider when purchasing a new vehicle. The top five factors identified were comfort (74%), cost (69.3%), fuel consumption (64.7%), engine performance (64%), and safety technology features (57.3%). These findings suggest that Malaysian drivers are attentive to comfort, affordability, efficient vehicle performance, and safety technologies in their purchasing decisions.

https://doi.org/10.31436/iiumej.v27i2.3968

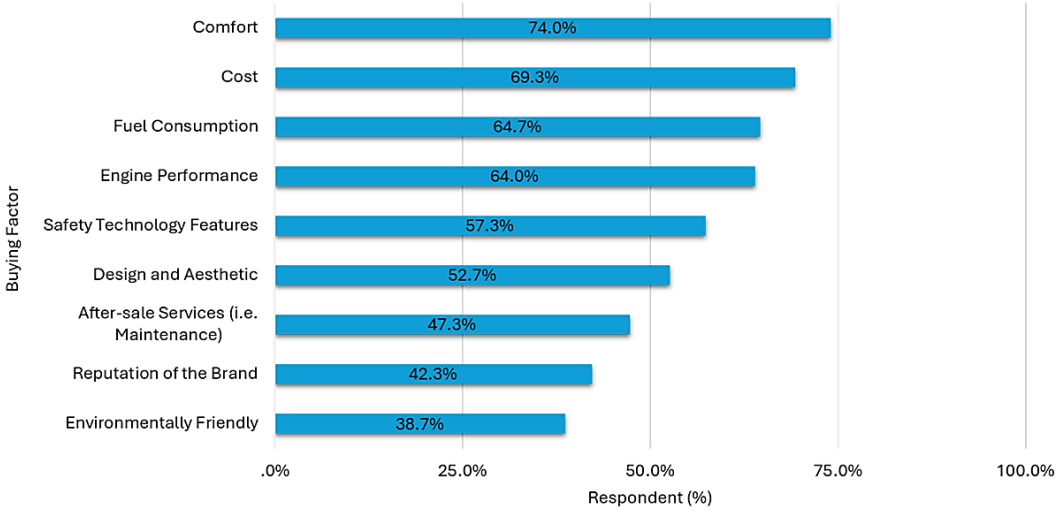


Figure 2. Contributing factors influencing vehicle purchase decisions.

Respondents were also asked about their awareness of existing in-vehicle safety technologies. Overall, 87% of respondents reported familiarity with these technologies, and 92.3% agreed that these systems are useful for preventing crashes, while 94.7% believed that the technologies are helpful during emergencies. However, several respondents admitted that they faced difficulties in differentiating the functions of the safety features and in using them correctly. Similar observations were reported by Kaur et al. [54], who reported that drivers with greater knowledge of vehicle safety systems tend to have stronger safety perceptions. Nonetheless, some drivers still appeared to underestimate these features, believing that their driving skills were sufficient. Conversely, Rohr et al. [55] noted that when drivers understand and use safety systems appropriately, fatalities can be reduced.

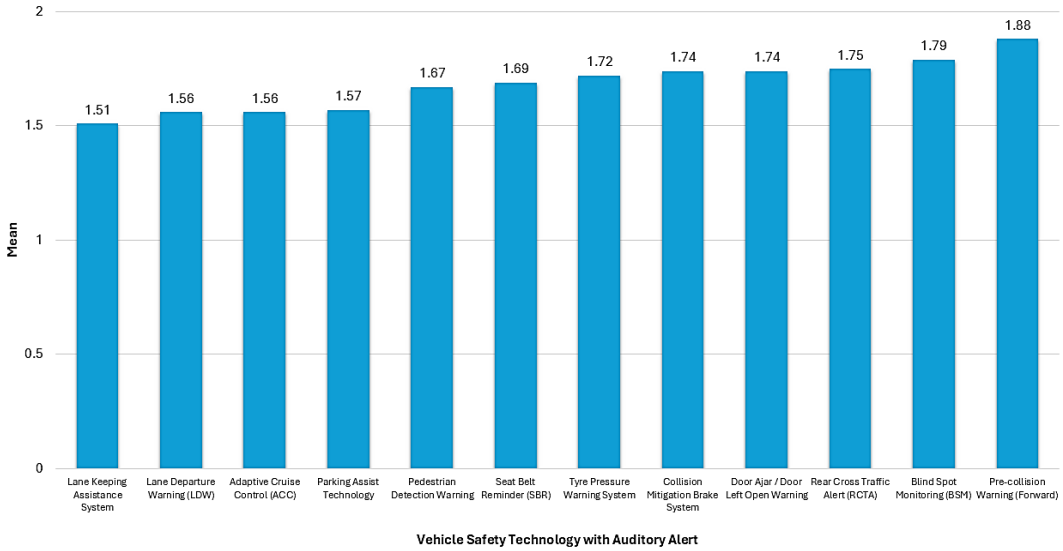


Figure 3. Importance of vehicle safety technology.

Figure 3 presents the perceived importance of 12 vehicle safety technologies among Malaysian drivers. These technologies are consistent with ASEAN NCAP's recommended safety requirements for vehicles in the Malaysian market. Respondents' awareness was evaluated using a three-point scale, where 2 indicated "very important," 1 "somewhat important," and 0 "not important." All technologies recorded mean scores above 1.5, indicating

https://doi.org/10.31436/iiumej.v27i2.3968

generally high awareness. From this study, it was found that Pre-Collision Warning (1.88, 88%) was rated as the most important, followed by Blind Spot Monitoring (1.79, 80%), Rear Cross Traffic Alert (1.75, 76%), Collision Mitigation Braking System (1.74, 75%), Door Ajar/Left-Open Warning (1.74, 76%), and Seat Belt Reminder (1.69, 72%). These results suggest that Malaysian drivers perceived collision-related and safety-reminder systems as the most important, reflecting their strong concern for crash prevention and emergency response.

In contrast, driving-control assistance technologies such as Lane Keeping Assistance System (LKAS) and Lane Departure Warning (LDW) were rated as less important, with mean scores of 1.51 and 1.56, respectively. These lower ratings may be attributed to less familiarity with semi-autonomous driving features. Overall, the findings indicate a strong presence of awareness among Malaysian drivers of vehicle safety technologies and their performance. This trend aligns with ASEAN NCAP and MIROS' ongoing efforts to mandate essential safety features in new vehicles, thereby increasing consumers' awareness of safety technologies. Variations in vehicle brand, cost, and safety features also appear to influence driver perceptions, with Malaysian drivers prioritizing systems that directly prevent crashes over those that require greater driver adaptation, such as Adaptive Cruise Control (ACC), Parking Assist, LKAS, and LDW.

3.2.3. Driver's Perception of Various Audible Alerts

Malaysian drivers' perceived ease of use of in-vehicle safety technologies is demonstrated in Figure 4. Overall, 76% of respondents agreed that these technologies are easy to operate. About 71.3% also found it easy to learn how the systems work, and 70.7% agreed that the interaction with the technology is clear and understandable. In addition, 62.7% felt that using the technologies requires minimal effort. These findings suggest that most Malaysian drivers perceive Advanced Driver Assistance Systems (ADAS) and related safety features as user-friendly. This outcome is consistent with the findings of Wu et al. [56], who emphasized that the effectiveness of such systems depends heavily on how well drivers understand and interact with them. Hence, ensuring a user-friendly design for in-vehicle safety technologies is essential for these technologies to function smoothly.

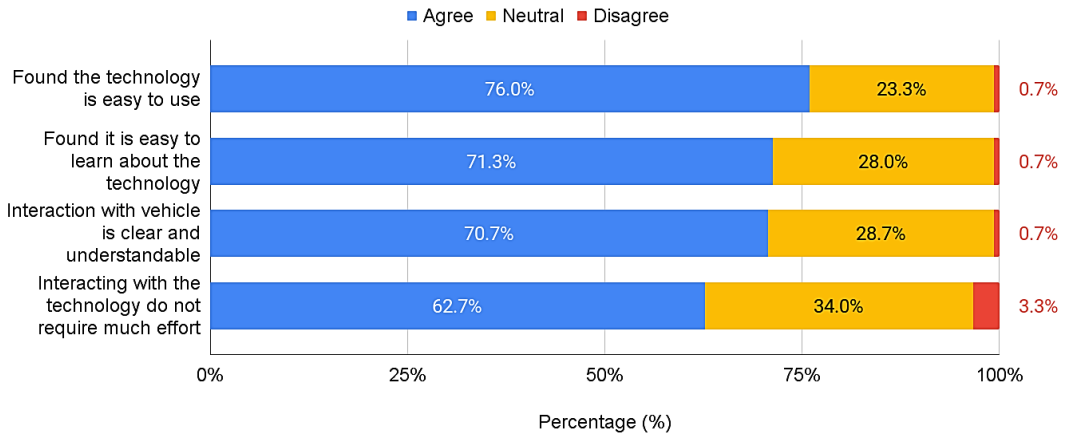


Figure 4. Perceived Ease of Use.

On the other hand, perceived usefulness reflects the extent to which drivers believe these technologies enhance their driving performance. As shown in Figure 5, more than 70% of respondents agreed that the technologies do assist them massively in driving and improving their overall driving performance, while 68% considered the technologies to be practically

https://doi.org/10.31436/iiumej.v27i2.3968

feasible. This, of course, increases the likelihood that drivers will willingly use the technologies. When drivers view a system as applicable, they are less likely to deactivate it.

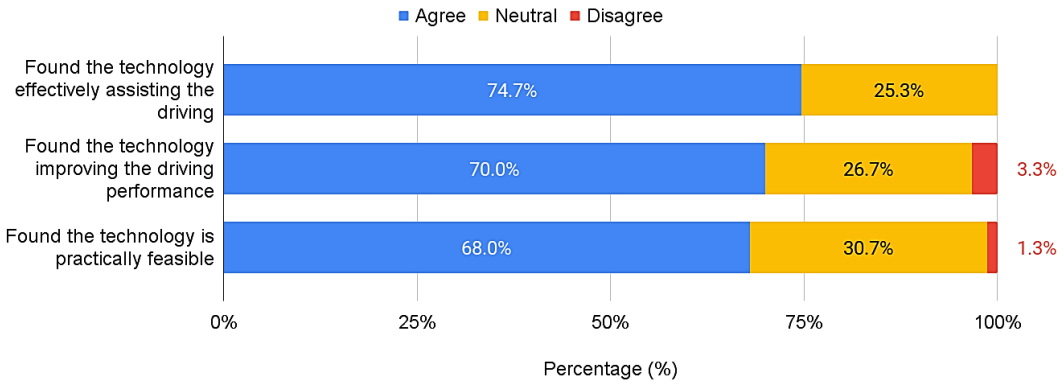


Figure 5. Perceived Usefulness.

The urgency level is defined as the effectiveness with which warnings can influence drivers' behavior in response to upcoming dangerous events. Figure 6 demonstrated that 86.7% of respondents believed that vehicle safety technology increases their driving safety and 68% trusted that the technology could reduce the likelihood of crashes. However, 48.7% of respondents found that the technologies distracted them while driving. The European Commission [57] study on driver distraction highlighted that vehicle safety systems are designed to draw the driver's attention away from other traffic activities. However, the warning alerts and signals must be instructive to the driver without being too overwhelming or causing cognitive distraction.

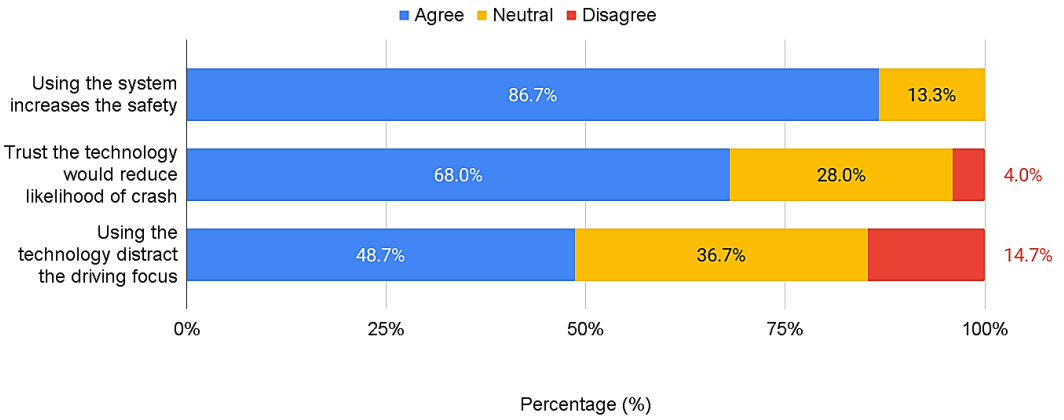


Figure 6. Perceived Safety.

In the audio test conducted in the survey, respondents selected their preferences for the perceived urgency of audio alerts, which encompassed three variables: sound pressure level, fundamental frequency, and an intermittent cycle. Prior to the audio test, the sound level meter was used to ensure that the ambient noise inside the car was 60 dB with the air conditioner and engine running. This study used four different sound pressure levels: 80 dB, 75 dB, 70 dB, and 65 dB. Respondents were asked to choose the preferred sound characteristic across different frequency settings and sound pressure levels. They were then required to rate their perceived annoyance and urgency for each frequency and intermittent cycle.

The audio used for the survey was set at three frequencies: 500 Hz, 1000 Hz, and 1500 Hz, while the intermittent cycle was set at three speeds: 500 ms, 1000 ms, and 1500 ms. The sound

https://doi.org/10.31436/iiumej.v27i2.3968

pressure level was set at 65 dB to achieve the desired level of perceived annoyance. Meanwhile, the sound pressure level was set to 65 dB and the frequency to 500 Hz, designed to enhance the perception of urgency. The results for the preferred frequency and sound pressure level among Malaysian drivers are illustrated in Figure 7, while the results for perceived annoyance and urgency across different frequencies and intermittent cycles are shown in Figure 8.

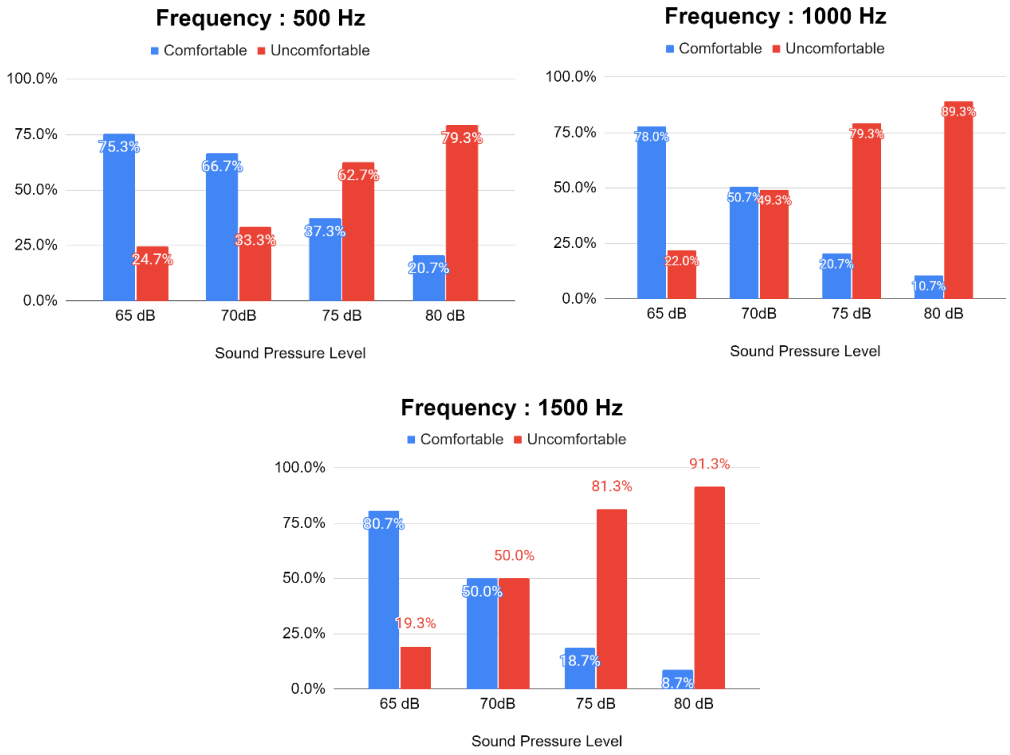


Figure 7. Preferred Sound Pressure Level and Frequency.

Figure 7 shows that respondents became increasingly uncomfortable as both the sound pressure level and frequency increased. At 500 Hz, 75.3% preferred 65 dB, while 79.3% found 80 dB uncomfortable. A small group, mainly aged 25–39 and 40–54, still tolerated 80 dB. Overall, 500 Hz tones were rated as least annoying (30%) and moderately urgent (60%). At 1000 Hz, 78% of respondents again preferred 65 dB as the most comfortable loudness, and this setting was perceived as fairly urgent and noticeably annoying. For 1500 Hz, 80.7% preferred 65 dB, whereas 91.3% judged 80 dB to be uncomfortable or noisy. Only 8.7% of drivers, mostly young drivers (18–24 and 25–39), tolerated 1500 Hz at 80 dB. Differences in comfort levels between younger and older drivers may reflect age-related changes in hearing sensitivity. Younger drivers are also more prone to external distractions, such as music, smartphones, and infotainment systems, as highlighted by the European Commission [57].

Compared with other frequency levels, audio at 1500 Hz was ranked the highest in perceived annoyance; yet 74.7% of respondents felt it was the most urgent (see Figure 8). Moreover, in terms of perceived annoyance and urgency across the different intermittent cycles, respondents rated the 500 ms cycle as the most irritating and the most urgent.

In terms of annoyance, more than 70% of respondents found 1000 ms and 2000 ms intermittent cycles not irritating; however, respondents perceived the 2000 ms cycle as the least urgent alert. Based on our findings, it can be concluded that Malaysian drivers prefer a frequency range of 1000-1500 Hz and a sound pressure level range of 65-70 dB as comfortable for auditory alerts. Furthermore, the 500 ms intermittent cycle is best suited for warning drivers

https://doi.org/10.31436/iiumej.v27i2.3968

of an urgent alarm, while the 1000 ms cycle is more suitable for recurring warning and reminder alerts.

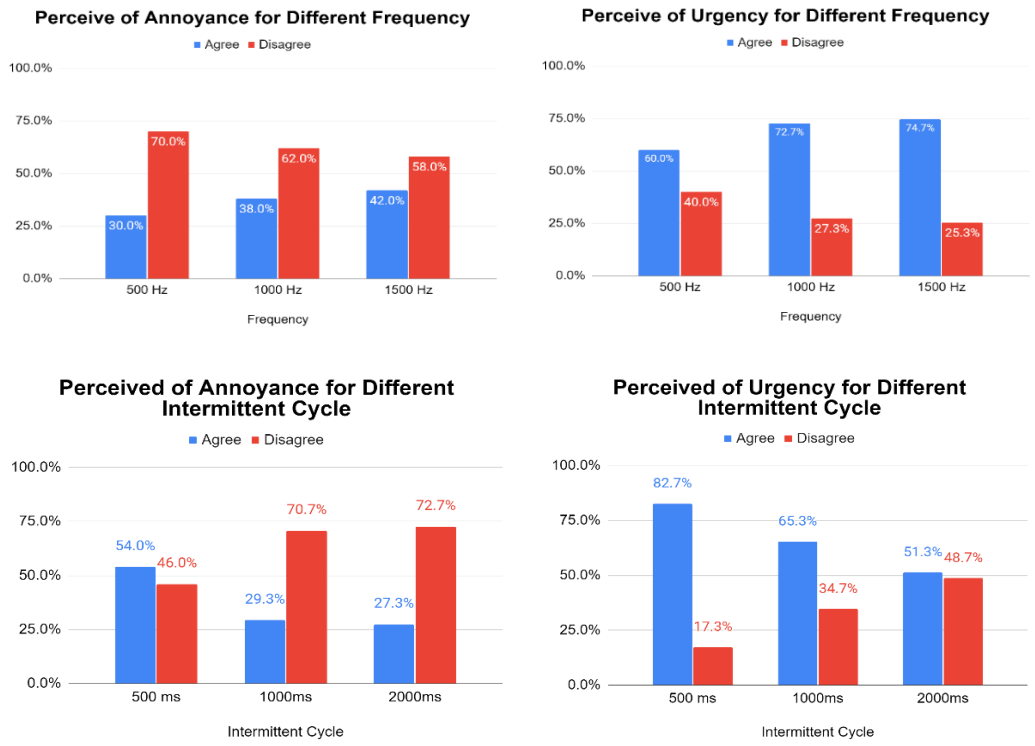


Figure 8. Perceived Annoyance and Urgency for Different Frequencies.

3.3. Phase 3: Optimum Sound Characteristic

Auditory modalities convey information to the driver, and different alert parameters correspond to different levels of urgency [58]. A driver's perception greatly influences how safety features are interpreted. As noted by Wu & Chan [59], the way information is processed can influence drivers' decision-making. This connection is also reflected in Malaysian consumers' tendency to consider ASEAN NCAP safety ratings when purchasing vehicles. However, as highlighted by Altamirano et al. [60] and other previous studies [19], Malaysian drivers have had limited exposure to Advanced Driver Assistance Systems (ADAS) due to their late introduction into the local market.

Figure 9 illustrates the frequency ranges used by five OEMs. Although most frequencies fall within the acceptable range defined in ISO 15006:2011, several exceed the recommended upper limit. This could potentially increase the likelihood of perceived distraction and annoyance among Malaysian drivers. The survey results further indicate that perceived urgency and comfort vary with frequency and sound pressure level. From the audio test, it was found that the chime configuration rated most favorably by Malaysian drivers was a 1000 Hz tone at 65 dB with an intermittent cycle of approximately 1000 ms. This configuration was perceived as sufficiently urgent for safety-critical scenarios, yet less annoying than higher-frequency tones. Therefore, this study recommends 1000 Hz at 65 dB as the preferred auditory-alert setting for in-vehicle safety systems in Malaysia.

https://doi.org/10.31436/iiumej.v27i2.3968

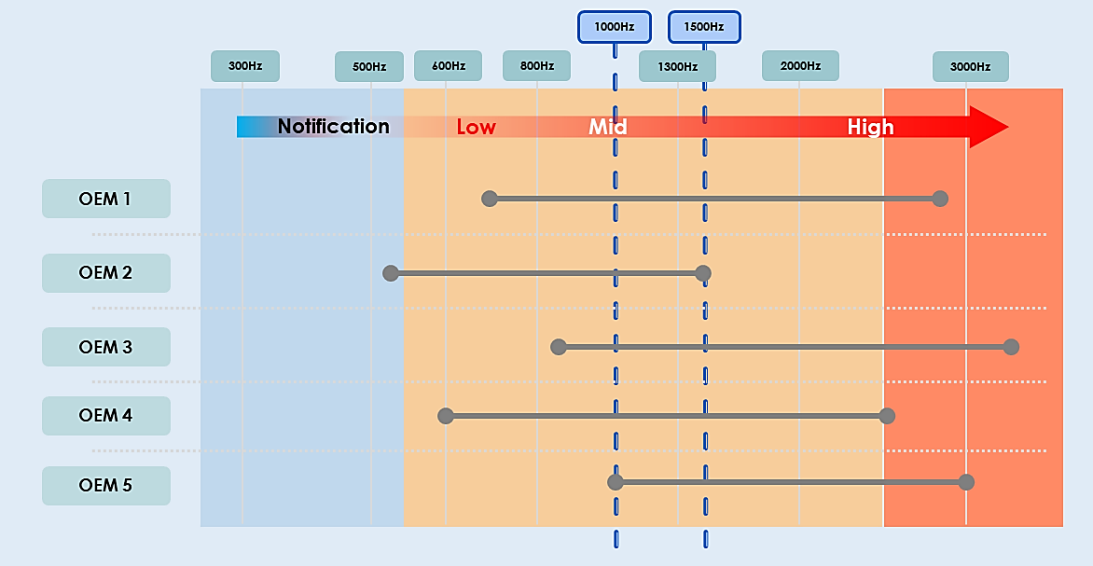


Figure 9. Range of frequency for vehicle safety technology and recommended frequency.

4. CONCLUSION

This study contributes deeper insight into Malaysian drivers' awareness, perceptions, and behavioral responses towards vehicle safety technologies that incorporate auditory alerts. Aside from identifying factors contributing to vehicle purchase decisions, the findings also highlight an overwhelming preference for safety-related technologies. Collision-avoidance safety technologies such as Pre-Collision Warning, Blind Spot Monitoring, Rear Cross Traffic Alert, Collision Mitigation Brake System, Seat Belt Reminder, and Door-Ajar Warnings were perceived as the most critical. On the other hand, lower familiarity with Lane Departure Warning and the Lane Keeping Assistance System reflects the limited awareness by Malaysian drivers of driving-control assistance technologies. This suggests a gap in public understanding of semi-autonomous safety features.

A key contribution of this research is to link driver perception to the technical properties of auditory alerts. The comparison between OEM-supplied frequencies and user preferences showed that although most OEM alerts fall within the ISO 15006:2011 guidelines, several exceed the optimal frequency range and may contribute to distraction or annoyance. The preferred auditory characteristics identified in this study, specifically 1000 Hz at 65 dB, provide an evidence-based reference for technology developers in designing alerts that balance urgency and comfort. These findings have direct implications for OEMs in calibrating warning tones to maintain safety relevance while minimizing distraction.

From a policy perspective, these findings carry implications for Malaysia's road-safety stakeholders. Firstly, ASEAN NCAP could incorporate auditory-alert preference data into the guidelines for local manufacturers. At the same time, OEMs may raise driver acceptance by tuning alert frequencies and sound pressure levels within user-preferred ranges. Furthermore, the Ministry of Transportation Malaysia and ASEAN NCAP may also consider standardizing minimum auditory-alert requirements for lower-specification vehicles to ensure consistent safety performance across vehicle models. Because prolonged exposure to excessively high frequencies can cause discomfort or hearing issues, it is essential to ensure that warning alerts balance urgency (to prompt a timely response) with minimal annoyance (to avoid distraction). Optimizing auditory-alert design in this way may support clearer human–technology interaction and contribute to safer driving experiences in Malaysia.

This study acknowledges certain limitations, particularly its Klang Valley-focused sample and the use of controlled stationary-vehicle testing. The involvement of participants from diverse regions and the investigation of cross-cultural differences in auditory perception should be explored in future research. It would also be beneficial to conduct simulation-based or real-world driving experiments to validate the effectiveness of alerts under dynamic driving conditions.

Despite these limitations, the findings offer meaningful insights for policymakers, manufacturers, and safety stakeholders. Designing auditory alerts that align with local driver preferences while still meeting international standards can enhance local driver acceptance, reduce distraction, and strengthen Malaysia's broader road-safety initiatives under the ASEAN NCAP protocol roadmaps and national safety plans.

ACKNOWLEDGEMENT

The authors would like to acknowledge the support and guidance provided by Universiti Teknologi Malaysia (UTM), grant number: Q.K130000.4443.4J766, the Malaysian Institute of Road Safety Research (MIROS), and the ASEAN NCAP secretariat through the ASEAN NCAP Collaborative Holistic Research (ANCHOR V) program (grant number: R.K130000.7314.1U087, R.K130000.7343.1U088). Special thanks to ACTS Smart Solutions Sdn. Bhd. for their assistance throughout the project. Greatest gratitude to all volunteers and respondents who are contributing to the completion of data collection.

REFERENCES

[1] Bassani M, Catani L, Hazoor A, Hoxha A, Lioi A, Portera A, Tefa L. (2023) Do driver monitoring technologies improve the driving behaviour of distracted drivers? A simulation study to assess the impact of an auditory driver distraction warning device on driving performance. Transportation Research Part F: Traffic Psychology and Behaviour, 95, 239–250. https://doi.org/10.1016/j.trf.2023.04.013

[2] Weibull K, Lidestam B, Prytz E. (2024) Optimising timing for emergency vehicle approaching warnings. Acta Polytechnica CTU Proceedings, 51, 125–128. https://doi.org/10.14311/APP.2024.51.0125

[3] Zou Z, Khan A, Lwin M, Alnajjar F, Mubin, O. (2025) Investigating the impacts of auditory and visual feedback in advanced driver assistance systems: a pilot study on driver behaviour and emotional response. Frontiers in Computer Science, 6, 1499165. https://doi.org/10.3389/fcomp.2024.1499165

[4] Rukonić L, Mwange MAP, Kieffer S. (2023) How older drivers perceive warning alerts? Insights for the design of driver–car interaction. SN Computer Science, 4(1), 56. https://doi.org/10.1007/s42979-022-01455-9

[5] Pada Lumba AF, Arifal Hidayat A, Ariyanto A, Harriad Akbar S. (2025) Probability of single-vehicle accidents among elderly motorcyclists in Indonesia. IIUM Engineering Journal, 26(3). https://doi.org/10.31436/iiumej.v26i3.3474

[6] Wen M, El Ali A, Jiang J, Ma R. (2022) Effect of spectral parameters on the elderly's urgency perception of auditory warning signals. Applied Acoustics, 194, 108798. https://doi.org/10.1016/j.apacoust.2022.108850

[7] Haas EC, Edworthy, J. (1996) Designing urgency into auditory warnings using pitch, speed, and loudness. Computing and Control Engineering Journal 1996, 7(4), 193–198. https://doi.org/10.1049/cce:1996040

[8] Baldwin CL, Lewis, BA. (2014) Perceived Urgency Mapping Across Modalities Within a Driving Context. Applied Ergonomics 2014, 45(5), 1270-1277. https://doi.org/10.1016/j.apergo.2013.05.002

[9] Wei D, Zhang C, Fan M, Ge S, Mi Z. (2024) Research on multimodal adaptive in-vehicle interface interaction design strategies for hearing-impaired drivers in fatigue driving scenarios. Sustainability, 16(24), 10984. https://doi.org/10.3390/su162410984

[10] Bliss JP, Gilson RD, Deaton JE. (1995) Human Probability Matching Behavior in Response to Alarms of Varying Reliability. Ergonomics 1995, 38, 2300–2312. https://doi.org/10.1080/00140139508925269

[11] Malhotra N, Shinar D. (2023) Cooperative warning systems: The impact of false and unnecessary alarms on drivers' compliance. Transportation Research Part F: Traffic Psychology and Behaviour, 92, 421-433. https://doi.org/10.1016/j.aap.2016.09.009

[12] World Health Organization. (2018) Global Status Report on Road Safety 2018: Summary (No. WHO/NMH/NVI/18.20). World Health Organization, 2018.

[13] United Nation. (2017) Road Safety Considerations in Support of the 2030 Agenda for Sustainable Development (UNCTAD/DTL/TLB/2017/4). United Nations Conference on Trade and Development.

[14] World Health Organization. (2011) Global Launch: Decade of Action for Road Safety 2011-2020 (No. WHO/NMH/VIP11. 08). World Health Organization.

[15] Ministry of Transportation Malaysia. (2021) Road Safety Plan (2021-2030). Available: https://www.mot.gov.my/en/land/safety/road-safety-plan-2021-2030.

[16] Hamid AA, Ishak NS, Roslan MF, Abdullah KH. (2023) Tackling human error in road crashes: An evidence-based review of causes and effective mitigation strategies. Journal of Metrics Studies and Social Science, 2(1), 1–9. Available: https://ejournal.papanda.org/index.php/jmsss/index

[17] Mohammed AA, Ambak K, Mosa AM, Syamsunur DA. (2019) Review of Traffic Accidents and Related Practices Worldwide. Open Transportation Journal 2019, 13(1). https://doi.org/10.2174/1874447801913010065

[18] ASEAN NCAP. (2020) ASEAN NCAP Labelling Guideline for Manufacturer (Malaysia Only) – Version 1.2.

[19] Syafiq M, Mohamed S. (2017) Malaysian Drivers' Perception Towards Advanced Driver Assistance Systems (ADAS). Human Factors and Ergonomics Journal 2017, 2(2), 15–20.

[20] Nees, MA, Walker BN. (2011) Auditory Displays for In-Vehicle Technologies. Reviews of Human Factors and Ergonomics 2011, 7(1), 58-99. https://doi.org/10.1177/1557234X11410396

[21] Spencea C, Driver J. (2017) Audiovisual Links in Attention: Implications for Interface Design. Engineering Psychology and Cognitive Ergonomics, Volume 2: Job Design and Product Design 2017, 7. https://doi.org/10.4324/9781315094489-24

[22] Moravčík Ľ, Jaśkiewicz M. (2018) Ways of Improving Car Safety in the EU Through Regulation. Perner's Contacts 2018, 13(2), 50-56.

[23] Huang A, Derakhshan S, Madrid-Carvajal J, Nosrat Nezami F, Wächter MA, Pipa G, König P. (2024) Enhancing Safety in Autonomous Vehicles: The Impact of Auditory and Visual Warning Signals on Driver Behavior and Situational Awareness. Vehicles, 6(3), 1613-1636. https://doi.org/10.3390/vehicles6030076

[24] Cahill CR, Gray GA, Klassen TR. (2019) Effect of Visual vs. Auditory Alerts on Older Drivers' Glances Toward Potential Hazards: An Examination of Effectiveness. Transportation Research Record: Journal of the Transportation Research Board, 2673(3), 118-127. https://doi.org/10.1177/0361198119844244

[25] Li L, Xu X, Wang X. (2024) Audiovisual messages may improve the processing of traffic information and vigilance in partially automated driving. Cognitive Research: Principles and Implications, 9(1), 25. https://doi.org/10.1186/s41235-024-00580-8

[26] Wang M, Liao Y, Lyckvi SL, Chen F. (2019) How Drivers Respond to Visual Vs. Auditory Information in Advisory Traffic Information Systems. Behaviour & Information Technology 2019, 1-12. https://doi.org/10.1080/0144929X.2019.1667439

[27] European Automobile Manufacturers' Association. (2023) ACEA regulatory guide 2023: Smarter regulation for a globally competitive European industry. European Automobile Manufacturers' Association. Available: https://www.acea.auto/files/ACEA-Regulatory-Guide-2023.pdf

[28] Edworthy J, Hellier E. (2006) Alarms and Human Behavior: Implications for Medical Alarms. British Journal of Anaesthesia 2006, 97, 12–17. https://doi.org/10.1093/bja/ael114

[29] Block FE, Nuutinen L, Ballast B. (1999) Optimisation of Alarms: A Study on Alarm Limits, Alarm Sounds, and False Alarms, Intended to Reduce the Annoyance. Journal of Clinical Monitoring and Computing 1999, 15, 75–83. https://doi.org/10.1023/A:1009992830942

[30] Sayed RA, Eskandarian A, Mortazavi A. (2012) Drowsy and Fatigued Driver Warning, Countermeasures, and Assistance. Handbook of Intelligent Vehicles 2012, 2, 977-990. https://doi.org/10.1007/978-0-85729-085-4_37

[31] JD Power (2019) U.S. Tech Experience Index (TXI) Study, 2019.

[32] National Highway Traffic Safety Administration (NHTSA). (2018) DOT HS 812 472: Vehicle Shopper's Guide - Driver Assistance Technologies. National Highway Traffic Safety Administration.

[33] Moravčík Ľ, Jaśkiewicz L. (2016) Integrated Intelligent Safety Systems. Perner's Contacts 2016, 11(2), 55-73. Available: https://pernerscontacts.upce.cz/index.php/perner/article/view/533

[34] Society of Automotive Engineers (SAE). (2015) SAE J3063 Standard: Active Safety Systems Terms & Definitions. Society of Automotive Engineering.

[35] Society of Automotive Engineers (SAE). (2020) SAE Recommended Common Naming for Advanced Driver Assistance Technologies. Society of Automotive Engineering.

[36] Mohd Zaki NI, Che Husin SM, Abu Husain MK, Abu Husain N, Ma'aram A, Marmin SNA, Adanan AF, Ahmad Y, Abu Kassim KA. (2020) Auditory Alert for In-Vehicle Safety Technologies: A Review. Presented at The 5th International Conference on Sustainable Mobility (ICSM2020), Kajang, Selangor, 2020. https://doi.org/10.56381/jsaem.v5i1.155

[37] Mohd Zaki NI, Che Husin SM, Abu Husain MK, Abu Husain N, Ma'aram A, Marmin SNA, Adanan AF, Ahmad Y, Abu Kassim KA. (2021) Auditory Alert for In-Vehicle Safety Technologies: A Review. Journal of the Society of Automotive Engineers Malaysia 2021, 5(1), 88-102, e-ISSN 2550-2239 & ISSN 2600-8092.

[38] International Organization for Standardization. (2011) ISO 15006:2011 Road Vehicles – Ergonomic Aspects of Transport Information and Control Systems - Specifications for In-Vehicle Auditory Presentation.

[39] Malaysian Automotive Association (MAA). (2015) Market Review for 2014 and Outlook for 2015. Malaysian Automotive Association.

[40] Malaysian Automotive Association (MAA). (2016) Market Review for 2015 and Outlook for 2016. Malaysian Automotive Association.

[41] Malaysian Automotive Association (MAA). (2017) Market Review for 2016 and Outlook for 2017. Malaysian Automotive Association.

[42] Malaysian Automotive Association (MAA). (2018) Market Review for 2017 and Outlook for 2018. Malaysian Automotive Association.

[43] Malaysian Automotive Association (MAA). (2019) Market Review for 2018 and Outlook for 2019. Malaysian Automotive Association.

[44] Malaysian Automotive Association (MAA). (2023) Market Review 2022. Malaysian Automotive Association.

[45] Malaysian Automotive Association (MAA). (2024) Market Review 2023. Malaysian Automotive Association.

[46] Ross KN. (2005) Quantitative Research Methods in Educational Planning: Module 3 – Sample Design for Educational Survey Research. International Institute for Educational Planning/UNESCO.

[47] Raosoft. (2020) Raosoft Sample Size Calculator. Raosoft, Inc., Seattle. Available: http://www.raosoft.com/samplesize.html. (Accessed on 14 September 2020)

[48] National Highway Traffic Safety Administration (NHTSA). (2013) Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices. NHTSA. Docket No. NHTSA-2010-0053.

[49] United Nations Economic Commission for Europe (UNECE). (2011) Guidelines on Establishing Requirements for High-Priority Warning Signals. World Forum for Harmonization of Vehicle Regulations.

[50] Roslin EN, Mohd Jawi Z, Abdul Jalal RI, Hadi AE, Daut AS, Hasan HMA, Ahamat MA, Ahmad Azmy NS, Abdul Jamal B. (2020) Malaysian Consumer Awareness and Perspectives towards ASEAN NCAP. Journal of the Society of Automotive Engineers Malaysia 2020, 4(1), 92–100. https://doi.org/10.56381/jsaem.v4i1.54

[51] Campbell JL, Brown JL, Graving JS, Richard CM, Lichty MG, Sanquist T, Morgan JL. (2016) Human Factors Design Guidance for Driver-Vehicle Interfaces (Report No. DOT HS 812 360). Washington, DC: National Highway Traffic Safety Administration. Available: https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/812360_humanfactorsdesignguidance.pdf

[52] Stevens A, Quimby A, Board A, Kersloot T, Burns P. (2002) Design Guidelines for Safety of In-Vehicle Information Systems. TRL Limited. Available: https://trimis.ec.europa.eu/system/files/project/documents/20060728_165141_88073_UG340_Final_Report.pdf

[53] Marshall D, Boyle LN, Wu X, Brown TL. (2018) Auditory Alert Characteristics Impact on Crash Avoidance Warning Response (No. DOT HS 812 511). United States. Department of Transportation. National Highway Traffic Safety Administration.

[54] Kaul V, Singh S, Rajagopalan K, Coury M. (2010) Consumer Attitudes and Perceptions about Safety and Their Preferences and Willingness to Pay for Safety. Report No. 2010-02-2336. SAE International, Warrendale.

[55] Rohr S, Lind R, Myers R, Bauson W, Koslak W, Huan Y. (2000) An Integrated Approach to Automotive Safety Systems. SAE Technical Paper, 2000-31-0346. http://doi.org/10.4271/2000-01-0346

[56] Wu X, Li C, Wang X, Xiang G, Deng H, Jiang Z, Peng Y. (2025) Research on drivers' hazard perception and visual characteristics before vehicle-to-powered two-wheeler collisions. Transportation Research Part F: Traffic Psychology and Behaviour. Advance Online Publication. https://doi.org/10.1016/j.trf.2025.06.014

[57] European Commission / ESRA. (2023) Thematic Report No. 3: Distraction and Fatigue. ESRA.

[58] Zhu A, Ma KH, Choi ATH, Hu D, Hu CP, Peng P, He J. (2025). The Effectiveness of Unimodal and Multimodal Warnings on Drivers' Response Time: A Meta-Analysis. Applied Sciences, 15(2), 527. https://doi.org/10.3390/app15020527

[59] Wu SI, Chan HJ. (2011) Perceived Service Quality and Self-Concept Influences on Consumer Attitude and Purchase Process: A Comparison Between Physical and Internet Channels. Total Quality Management 2011, 22(1), 43–62.

[60] Altamirano A, Paredes R. (2022) Willingness to Pay for Advanced Safety Features in Vehicles: Evidence from Stated Preference Surveys. Transportation Research Record: Journal of the Transportation Research Board, 2676(3), 284-295. https://doi.org/10.1177/03611981221077077

https://doi.org/10.31436/iiumej.v27i2.3971

Field and Laboratory Evaluation of Reliability in a Novel Cone Deployment System for Safer Highway Operations

AZILAH ISMAIL[1, 2], MOHD AZLI SALIM[*2], NOR AZMMI MASRIPAN[2], MOHD ZAID AKOP[2], SITI WARDATULAINA MOHD YUSOF[3], CHONLATEE PHOTONG[4], ALAN WATSON[5], MOHAMMED HUSSIN A. AL-MOLA[6]

[1]Jabatan Perdagangan, Politeknik Ungku Omar, Ipoh, Perak, Malaysia
[2]Fakulti Teknologi dan Kejuruteraan Mekanikal, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia
[3]Fakulti Pengurusan Teknologi dan Teknousahawan, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia
[4]Faculty of Engineering, Mahasarakham University, Thailand
[5]Power Electronics and Machines Centre, The University of Nottingham, UK
[6]College of Agriculture and Forestry, University of Mosul, Iraq

**Corresponding author: azli@utem.edu.my*

(Received: 3 October 2025; Accepted: 29 January 2026; Published online: 10 May 2026)

ABSTRACT: This study evaluates the reliability and maintainability of the Cone Collecting and Laying (C2L) machine, developed to enhance safety and efficiency in highway maintenance operations by automating the deployment of traffic cones. The objective is to investigate the use of reliability metrics, which are Mean Time to Failure (MTTF), Mean Time to Repair (MTTR), and Mean Time Between Failures (MTBF), as a systematic approach to assess the operational efficiency and failure behavior of the C2L machine in real highway maintenance scenarios. The methodology encompasses controlled laboratory and field testing conducted over 68 hours in Test Motion A and 3,192 hours in Test Motion B. During these tests, real-time failure data were recorded for six critical components of the C2L machine, namely the bearing, cone lifter, chain, sensor, PLC, and hook spring. Findings from Test Motion A revealed component vulnerabilities, including a low MTTF in bearings and a high MTTR in welded joints. In contrast, Test Motion B identified frequent failures in the PLC system, attributed to vibration stress, as well as a rare but critical hook spring failure caused by material fatigue. All reliability metrics were calculated from empirical failure and repair times, enabling accurate quantification of component durability and overall system availability. Failure data were further categorized into repairable and non-repairable types to inform targeted design improvements. In conclusion, integrating MTTF, MTTR, and MTBF into the evaluation of the C2L machine proved essential for identifying weaknesses and enhancing operational performance. The study recommends proactive maintenance scheduling, structural reinforcement, and redesign of vulnerable components to ensure system resilience. These findings contribute to improved worker safety, reduced operational downtime, and the enhanced long-term deployment potential of the C2L machine in highway maintenance applications.

ABSTRAK: Kajian ini menilai kebolehpercayaan dan kebolehselenggaraan mesin Pemungut dan Meletak Kon (C2L) bagi meningkatkan keselamatan dan kecekapan operasi penyelenggaraan lebuh raya melalui automasi proses meletak dan mengutip kon trafik. Objektif kajian ini adalah meneliti penggunaan metrik kebolehpercayaan seperti Masa Purata Sebelum Kegagalan (MTTF), Masa Purata Membaiki (MTTR) dan Masa Purata Antara Kegagalan (MTBF) sebagai satu pendekatan sistematik bagi menilai kecekapan operasi serta tingkah laku kegagalan mesin C2L dalam situasi penyelenggaraan lebuh raya sebenar. Kaedah

https://doi.org/10.31436/iiumej.v27i2.3971

ini merangkumi ujian makmal dan ujian lapangan terkawal yang dijalankan selama 68 jam bagi Ujian Gerakan A dan 3,192 jam bagi Ujian Gerakan B. Sepanjang ujian tersebut, data kegagalan masa nyata direkodkan pada enam komponen kritikal mesin C2L, iaitu pengalas, pengangkat kon, rantai, pengesan, PLC dan spring cangkuk. Penemuan daripada Ujian Gerakan A menunjukkan kelemahan komponen termasuk MTTF yang rendah pada galas serta MTTR yang tinggi pada sambungan kimpalan. Sebaliknya, Ujian Gerakan B mengenal pasti kegagalan kerap pada sistem PLC akibat tekanan getaran, serta kegagalan kurang tetapi kritikal pada spring cangkuk disebabkan oleh kelemahan bahan. Semua metrik kebolehpercayaan dikira daripada data empirikal masa kegagalan dan masa pembaikan, sekaligus membolehkan pengukuran ketahanan komponen serta ketersediaan keseluruhan sistem. Data kegagalan turut dikategori pada jenis boleh dibaiki dan tidak boleh dibaiki bagi menyokong penambahbaikan reka bentuk terfokus. Kesimpulan, integrasi MTTF, MTTR dan MTBF dalam penilaian mesin C2L terbukti penting bagi mengenal pasti kelemahan serta meningkatkan prestasi operasi. Kajian ini mengesyorkan jadual penyelenggaraan proaktif, pengukuhan struktur, dan reka bentuk semula komponen yang lemah bagi memastikan daya tahan sistem. Dapatan ini menyumbang kepada peningkatan keselamatan pekerja, pengurangan masa henti operasi, serta potensi penggunaan jangka panjang mesin C2L dalam aplikasi penyelenggaraan lebuh raya.

KEYWORDS: *C2L Machine, Failure Analysis, Reliability Metrics, MTTF, MTTR, MTBF.*

1. INTRODUCTION

Highway maintenance operations frequently involve manual handling of traffic cones, a task that poses significant safety risks to workers and contributes to physical fatigue. By addressing these issues, the Cone Collecting and Laying (C2L) machine was developed as a semi-automated solution capable of deploying and retrieving cones from a moving lorry. While this innovation improves operational safety and efficiency, ensuring the long-term reliability and maintainability of the C2L machine is essential for its successful integration into routine highway maintenance activities. This requires a systematic evaluation of its key components under realistic, simulated operational conditions.

Then, by quantifying the reliability and availability of the C2L machine, this study employs three fundamental reliability metrics: Mean Time to Failure (MTTF), Mean Time to Repair (MTTR), and Mean Time Between Failures (MTBF). These time-based indicators provide important information on component durability, specifically, how long components operate before failing (MTTF), how quickly they can be repaired (MTTR), and the average duration between successive failures (MTBF). These metrics help determine whether the machine performs well during long highway operations and identify components that often fail or take a long time to fix. MTTF indicates how long a non-repairable part can operate before it fails. MTTR shows how long it takes to repair a part after it fails. MTBF combines both values to show the average time between failures for components that can be repaired. Therefore, by applying these metrics, the study aims to identify failure components of the C2L machine in real highway maintenance scenarios.

Therefore, to achieve this objective, comprehensive performance evaluations were conducted under two testing conditions: smooth (Test Motion A) and bumpy (Test Motion B) road surfaces. These tests assess how each component (bearings, welded joints, sensors, PLC units, chains, and hook springs) performs, responds to stress, and fails under different operating conditions, providing a complete understanding of the machine's reliability in real-world highway maintenance scenarios. MTTR, MTTF, and MTBF are widely recognized within engineering literature as critical metrics for evaluating system performance and availability [1].

The MTTR represents the average time required to detect, diagnose, repair, and restore a system or component to operational status, covering all stages from fault notification to post-repair testing.

Reducing MTTR is important because shorter repair durations directly improve availability and minimize downtime. Meanwhile, MTTF estimates the expected operational lifespan of non-repairable components [2], while MTBF reflects the average time between failures for repairable components and is a primary measure of overall system reliability [3]. These metrics are used to identify common failure modes, evaluate system robustness, and support predictive maintenance strategies. Moreover, Markov chain analysis has proven useful in modeling complex system reliability, offering valuable insights into performance prediction and maintenance scheduling [4].

Despite the importance of these metrics for assessing machinery reliability, a review of the existing literature reveals a notable research gap. Most studies on traffic cone deployment systems have focused on automation mechanisms, system design, or safety improvements, with little to no emphasis on thorough reliability analysis. Previous research lacks empirical investigations into the failure behavior and maintainability of such machines under realistic operational conditions. The study by Mehdi et al. is closely related because it highlights the importance of comprehensive reliability, availability, and maintainability (RAM) analysis in modern mechatronic systems [5]. It also demonstrates that such detailed analyses are still rarely applied in many domains of applied machinery, including traffic cone deployment systems. This gap highlights the need for a detailed, metric-driven evaluation of system durability and failure trends, an area this study directly addresses.

In the context of the C2L machine, these reliability metrics are especially applicable. High MTTF and MTBF values indicate that components can operate longer, uninterrupted, thereby improving efficiency. Conversely, a low MTTR suggests rapid recovery from failures, ensuring minimal downtime. As emphasized by [6], strategies to maximize MTBF (or MTTF) and minimize MTTR are important for achieving optimal system availability. Additionally, [7] notes that systems engineering practices incorporate these reliability and maintainability indicators within broader frameworks of system effectiveness. When combined, these metrics offer a holistic view of a machine's dependability and provide the data needed for evidence-based decision-making in maintenance planning and design refinement [8]. Therefore, by employing MTTR, MTTF, and MTBF in its research methodology, this study facilitates the systematic evaluation and continuous performance enhancement of the C2L machine. The results not only diagnose component-level vulnerabilities but also drive targeted design improvements. High MTTF and MTBF correlate with fewer disruptions, while low MTTR enables efficient maintenance response. The analysis of both repairable and non-repairable failures thus builds a strong case for design optimization. In turn, this study highlights the value of reliability engineering in advancing automation technologies for highway maintenance, improving worker safety, reducing downtime, and enhancing overall system sustainability.

2. METHODOLOGY

2.1. Determining reliability metrics from C2L machine failure data

An evaluation of the operational reliability and maintainability of the C2L machine was conducted through two structured field tests that replicated real-world highway maintenance scenarios. The C2L body serves as the main structural framework of the C2L machine, supporting and connecting all major components, including the cone-laying mechanism, control systems, and supporting elements. Built for strength and durability, the C2L body

https://doi.org/10.31436/iiumej.v27i2.3971

enables the machine to operate safely and efficiently under the demanding conditions of highway maintenance. Its structurally reinforced design contributes significantly to the system's overall reliability and performance. Figure 1 presents the key components of the C2L machine, showing both the front and back views. Each component supports the automated functions of laying and collecting traffic cones.



(a) Front view (b) Back view

Figure 1. C2L body.

Figure 2 shows the complete design of the laboratory test rig used to evaluate the C2L machine. The rig functions as a controlled platform that safely replicates real road conditions while enabling consistent test repetition. Its structure features a long, sturdy frame with parallel guiding rails that enable smooth, stable operation of the C2L machine. A motorized unit is installed at one end to drive or simulate the machine's motion along the rails. At the center of the rig, a supporting frame accommodates the tipping mechanism, guiding posts, and any other attachments required to assess the cone collection and laying functions.

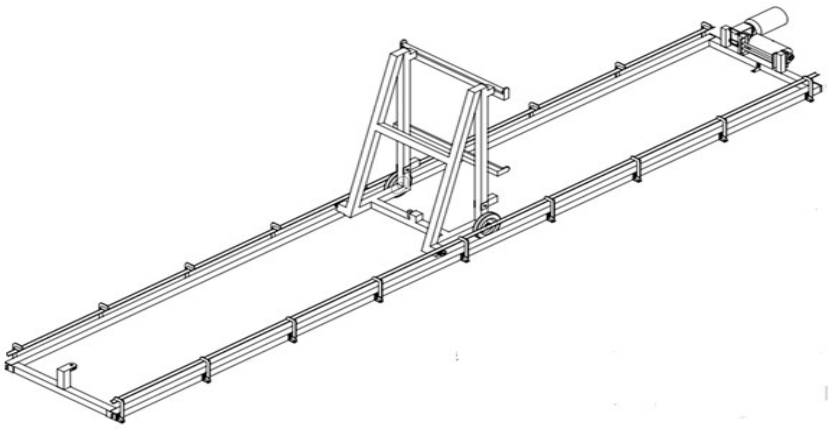


Figure 2. C2L machine laboratory test rig configuration.

Test Motion A examined the machine's performance over 68 hours on smooth road surfaces, while Test Motion B extended the evaluation to 3,192 hours on uneven, bumpy terrain. Each test is carried out in 12 sequences. Conducting 12 sequences for both motions ensured sufficient data for accurate, statistically meaningful analysis while keeping the testing process efficient in terms of time and resources. This approach enabled a thorough assessment of the C2L machine's performance, reliability, and operational consistency under both smooth and uneven surface conditions. For every sequence, the total testing time and the number of completed cycles were carefully recorded. These contrasting environments provided important information about the machine's performance under different stress conditions.

Reliability metrics, MTTF, MTTR, and MTBF were applied to assess the endurance and maintainability of six critical components: bearings, chains, sensors, cone lifter, PLC, and hook springs. These components were selected based on their significance to the machine's main operations. Table 1 summarizes the role of each part and presents quantitative data highlighting their durability, ease of maintenance, and contribution to overall system availability. Such measurements are essential in ensuring the C2L machine can function reliably in demanding roadside environments.

The primary aim of this methodology was to assess whether the C2L machine could effectively replace manual cone deployment and retrieval, thus enhancing safety and operational efficiency during highway maintenance. Reliability, as described by [9, 10], represents the likelihood that a component or system performs its intended function without failure over a defined period under specific conditions. In this study, MTTF, MTTR, and MTBF served as key performance indicators. Their application enables a clear understanding of both the frequency of failure and the effectiveness of repair processes. These findings support data-informed decisions to improve component design, refine maintenance strategies, and enhance the long-term dependability of the C2L machine in real-world operations, as stated in Eq. (1).

$$\begin{gathered} MTTR = Total\ Repair\ Time - Number\ of\ Failures \\ MTTF = Total\ Operating\ Time - Number\ of\ Failures \\ MTBF = MTTF + MTTR \end{gathered} \tag{1}$$

These time-dependent metrics require accurate data collection under varied operational conditions to yield meaningful findings. Such a comprehensive approach enables accurate evaluation of component durability, identification of design vulnerabilities or structural weaknesses, and development of targeted strategies for system improvement. MTTF quantifies the average time-to-failure of non-repairable components. This measure is essential for understanding component reliability and supports the implementation of proactive maintenance schedules. Therefore, by identifying potential failures, system availability can be improved, and operational disruptions during highway maintenance activities can be significantly reduced.

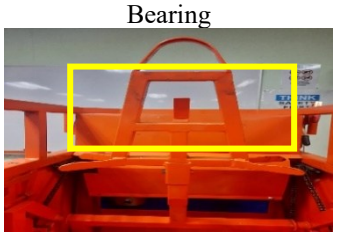
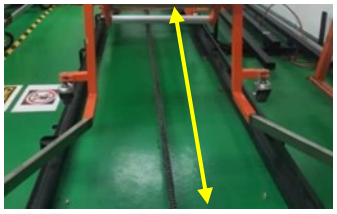
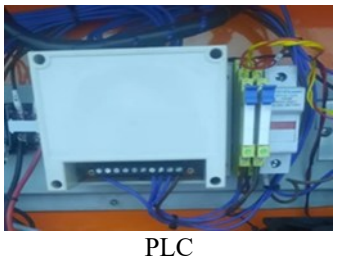
The MTTR quantifies the average time required to detect, diagnose, repair, and verify a failed component before restoring it to full functionality. Minimizing MTTR is crucial for reducing system downtime and maximizing availability. As highlighted by [11], effective management of MTTR directly impacts operational continuity. Maintenance engineering frameworks often incorporate reliability metrics like MTTR to optimize lifecycle costs [12].

Meanwhile, the MTBF measures the average interval between successive failures of repairable components, serving as an important indicator of overall system reliability and guiding maintenance prioritization [13]. Additionally, [14] emphasizes that proactive maintenance substantially prolongs machine lifespan in industrial environments. In applying

https://doi.org/10.31436/iiumej.v27i2.3971

these metrics to the C2L machine, systematic data collection was carried out during laboratory and field tests [15]. MTTF was calculated for non-repairable components such as sensors and motors, MTTR was derived from recorded repair durations following failures, and MTBF reflected the operational intervals between failures.

Table 1. The function of the component C2L machine in providing quantifiable data

Component	Function of Component C2L Machine
 Bearing	Roller bearings use cylindrical rollers to minimize friction and support rotating shafts or axles. They consist of cylindrical rollers arranged between an inner and an outer ring.
 Cone lifter	The cone lifter is designed to guide the cone during lifting, apply controlled force for stable handling, offer a defined path for cone movement, and ensure smooth lifting through its mechanical assembly.
 Chain	The chain consists of a series of interlinked metal segments positioned along the conveyor belt's centerline.
 Sensor	A motion sensor is a device that detects and reacts to particular environmental factors, such as light, heat, motion, moisture, and pressure. The sensors detect the C2L machine's position and send signals to control the motor's direction, ensuring it moves forward and backward along the test rig as required.
 PLC	The PLC accurately manages various components of mechanical equipment, such as motors, valves, and sensors. The damper box is a protective enclosure designed to safeguard the PLC from environmental and mechanical hazards.
 Hook spring	The coil hook spring is used as a connector linking the chain to the load, functioning as an integral part of the cone lifter. The hook spring helps maintain proper alignment and movement by absorbing forces, managing tension, and ensuring smooth operation.

This comprehensive reliability assessment identified mechanical weaknesses, including fatigue in welded joints and limitations within control systems, which informed targeted design

https://doi.org/10.31436/iiumej.v27i2.3971

improvements such as reinforced welds and upgraded control units [16]. By integrating these reliability metrics into the methodology, the study adopts a data-driven approach to enhancing the C2L machine's operational dependability [17]. This not only improves reliability and availability but also contributes to safer, more efficient highway maintenance operations.

Reliability, denoted as $R(t)$, is defined as the probability that a system will perform its intended function without failure under specified conditions over a given time [18]. Typically, reliability follows an exponential failure distribution expressed as:

$$R(t) = e^{-\lambda t} \tag{2}$$

where λ represents a constant failure rate, and t denotes time.

This model reflects an exponential decline in reliability over time, with maximum dependability at the start of operation [19, 20]. In contrast, availability measures the likelihood that a system is operational at any given moment, considering repair times, and is mathematically defined as:

$$A = \frac{MTBF}{MTBF+MTTR} \tag{3}$$

Here, $MTBF$ and $MTTR$ are key parameters derived from empirical testing, industry benchmarks, or expert evaluations. Besides, low reliability leads to frequent unplanned stoppages, reducing overall system availability by increasing downtime and operational interruptions [21-23].

2.2. Performance measurement and reliability assessment methodologies

The laboratory experiment was conducted in a controlled environment, with the room temperature consistently maintained at an average of 24°C to ensure stable testing conditions. The test rig, illustrated in Figure 3, was assembled in a secure area at the rear of the laboratory, with barricades installed to safeguard other users. The setup comprised several key components, including a DC motor, a roller, a chain, a sensor, a ramp, a cone lifter, and a tipping bar.

Although the laboratory test was conducted at an average room temperature of 24°C, the higher ambient temperatures typically experienced on highways, which can reach up to 45°C, are not expected to affect the C2L machine's cone collecting and laying functions. This is because its key mechanical and electrical components are engineered and rated to operate reliably across a broader temperature range. Therefore, the performance and motion behaviors observed at 24°C can be confidently expected to remain stable under higher real temperatures. During testing, data were recorded each time the C2L machine completed a full cycle, traveling from the test rig's starting point to the end and then returning to its original position.

This approach is consistent with the methodology employed by Lacarella et al. [24], which similarly regulates room temperatures during laboratory experiments to simulate natural stream conditions, thereby relating their findings to ecosystem restoration strategies aimed at mitigating temperature-related effects. In this study, the C2L machine was tested under dynamic operation by repeatedly moving between two motion sensors installed on a specially designed test rig. During each pass, the machine executed its designated cone pick-and-lay mechanism. The data were systematically recorded only after the completion of each full cycle, defined as movement from the starting point to the far end of the rig and back. Therefore, by preserving data integrity, only fully completed cycles were included in the analysis, except for static test runs specifically focused on evaluating the ramp's vertical movement.

https://doi.org/10.31436/iiumej.v27i2.3971

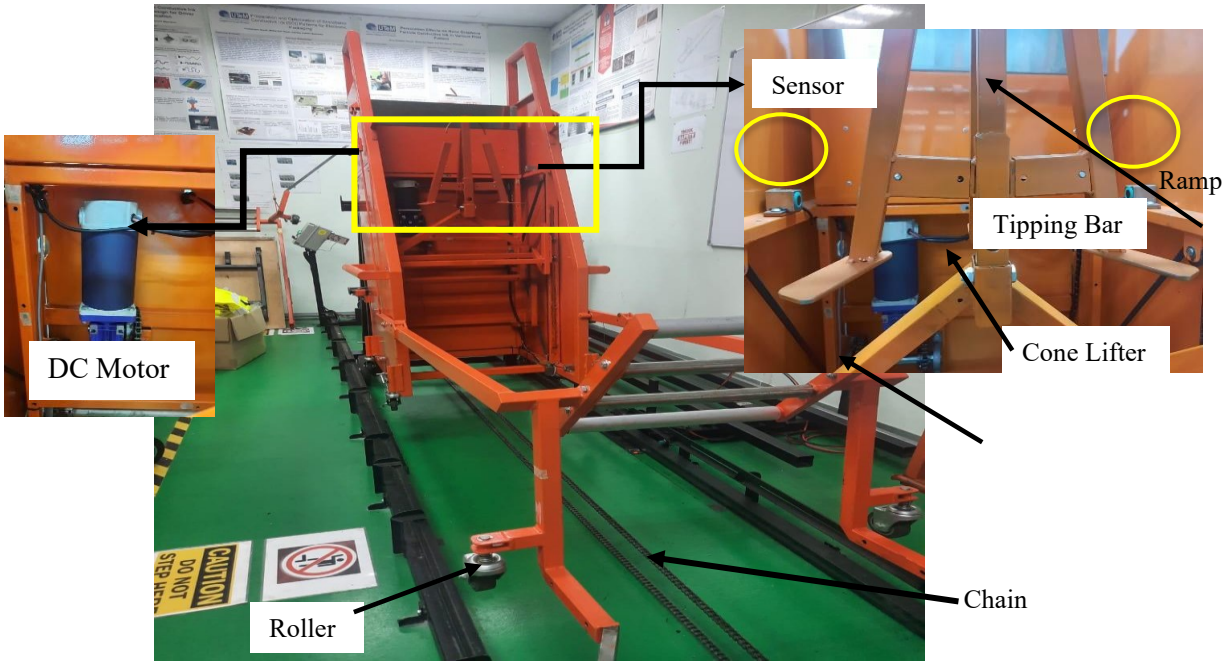


Figure 3. Test rig configuration in the laboratory.

The experimental schedule for both Test Motion A and Test Motion B spanned 14 weeks. All testing was conducted during regular working hours, with a maximum runtime of eight hours per day, five days a week. These operational constraints ensured the testing was carried out under close supervision, thereby guaranteeing the reliability and safety of the experiments. Following the conclusion of Test Motion A and Test Motion B, data collection concentrated on key parameters to evaluate the C2L machine's reliability and to identify potential design improvements. One primary metric recorded was the number of hours the machine operated daily during testing. A counter installed within the control box tracked the total number of completed runs. Throughout the testing phase, various performance metrics were recorded, including MTTF, MTBF, MTTR, Reliability Function (RF), and Non-Reliability Function (NRF). On the other hand, Żyluk et al. [25] advocate the use of the MTTF indicator as a strategic tool to enhance logistical support management within operational processes, optimize maintenance scheduling, and improve system reliability. These methodologies draw upon the approaches detailed in the references.

All experimental data, including recorded failures, were systematically compiled and analyzed. Tests were performed under varying conditions and repeated multiple times to ensure accuracy and reliability. In the event of failure, repairs were promptly conducted before resuming tests. By further refining the design and eliminating NRFs, the TRIZ methodology was applied to resolve technical contradictions and enhance overall system efficiency. Ultimately, the reliability analysis assessed the C2L machine's performance by evaluating the strength and durability of its components throughout operational use.

3. Results and Discussion

3.1. Component failure analysis from Test Motion A

The C2L machine encountered four critical failures during 68 hours of testing under smooth road conditions (Test Motion A). The bearing failure, as detailed in Table 2 (15 minutes' repair), was most likely caused by friction-induced wear resulting from repeated

https://doi.org/10.31436/iiumej.v27i2.3971

bidirectional rail movement. This issue occurred because the bearing had a low thread count, which weakened its ability to remain securely fixed under repetitive motion. This form of wear can lead to increased play within the bearing housing, elevating the risk of misalignment and subsequent operational jamming. Therefore, to enhance durability, it is recommended to use sealed, high-load-capacity bearings with improved lubrication retention, along with regular condition monitoring via vibration or thermal analysis. The welded joint failure, located below the cone lifter, which cracked due to fatigue, required the longest repair time (60 minutes) and revealed a significant structural vulnerability. The repeated cyclic stresses induced by operational vibration may have led to fatigue cracking at weld toes, common initiation points for structural failure. Employing full-penetration welds, improving joint geometry to distribute stress more evenly, and incorporating non-destructive testing (NDT) methods such as ultrasonic inspection during fabrication can significantly bolster reliability in these areas.

Table 2. Summary of failure components and repair data for the C2L machine during Test Motion A

Component	Total failures (count)	Total Repairs (count)	Repair Time (minutes)	Total testing time (hours)
Bearing	1	1	15	68
Welded Joint	1	1	60	68
Chain	1	1	40	68
Sensor	1	1	30	68

The chain failure (40 minutes' repair) dislodged from the gear during the test was attributed to tensile overload during cone-laying cycles, which may have caused elongation or breakage of links, in turn risking damage to neighboring drivetrain components. Enhancing chain durability could involve selecting high-tensile alloy chains with wear-resistant coatings, ensuring correct tensioning, and introducing guided sprockets to minimize lateral movement and misalignment. The sensor failure (30 minutes' repair) occurred when it was unable to generate the signal needed to continue the motion cycle, causing the operation to stop. This malfunction was likely due to either an electrical fault or mechanical stress from environmental vibrations. MTTF, MMTR, and MTBF provide important information on component durability, maintainability, and overall system reliability, which are essential for ensuring safe and efficient highway maintenance operations.

Firstly, the MTTF quantifies the expected service life of non-repairable components such as sensors and welded joints. In Test Motion A, all components had an MTTF of 68 hours, calculated as the total test duration divided by the number of failures per component. This exceptionally low MTTF, especially for the welded joint, indicates a structural weakness under cyclic loading conditions. Such failures are often attributed to fatigue-induced microcracking at stress-concentration points within the weld. Enhancing MTTF may involve material upgrades, such as switching to high-grade structural steel with superior fatigue resistance, or implementing design changes inspired by TRIZ, for instance, replacing welded connections with bolted joints that enable modular disassembly and more uniform stress distribution. Increasing MTTF directly reduces the likelihood of premature in-service failures, thus improving operational safety and minimizing unplanned maintenance costs [26].

Secondly, the MTTR assesses the ease and speed with which failed components can be restored to working condition. The welded joint exhibited the highest MTTR of 60 minutes, reflecting the complexity of structural repairs, which often involve grinding, re-welding, and post-weld inspection. Long MTTRs negatively impact system availability. Reducing MTTR, such as redesigning components for easier access or adopting quick-release fasteners,

https://doi.org/10.31436/iiumej.v27i2.3971

significantly improves maintainability. For example, halving the MTTR for sensor replacements (from 30 to 15 minutes) could increase overall system availability by approximately 8%, ensuring the machine can resume highway operations with minimal downtime.

Thirdly, MTBF applies to repairable components such as chains and bearings. It combines MTTF and MTTR to represent the average interval between failures. The chain's MTBF of 68.67 hours suggests frequent operational disruptions, potentially due to inadequate lubrication, misalignment, or material fatigue. Implementing scheduled lubrication regimes, using corrosion-resistant alloy chains, or installing chain tensioners can significantly enhance MTBF. A 20% improvement in MTBF not only extends service intervals but can also reduce spare part inventory by up to 15%, lowering lifecycle costs. When MTBF increases, the number of breakdowns over a given period naturally decreases. As a result, maintenance teams do not need to replace components as frequently, which directly affects inventory planning. For consistent field performance, an MTBF target exceeding 100 hours is advisable to ensure uninterrupted deployment over extended highway maintenance shifts. Therefore, improving these reliability metrics through targeted design, maintenance planning, and material selection will substantially enhance the C2L machine's dependability, reduce operational interruptions, and promote long-term cost efficiency.

3.2. Reliability metrics for failed components in Test Motion A

The analysis of reliability metrics for the C2L machine during Test Motion A provides important insights into the performance, durability, and failure characteristics of its major components. The evaluation, based on MTTF, MTTR, and MTBF, focuses on four critical parts: the bearing, welded joint, chain, and sensor. The results highlight key vulnerabilities and focus on the importance of proactive maintenance strategies, informed design improvements, and enhanced diagnostic practices to support long-term field deployment.

The bearings recorded in Table 3 show the lowest MTTF (2.25 hours), indicating a higher likelihood of early failure compared to the other components. Although the failure was resolved quickly with an MTTR of just 15 minutes, the early breakdown emphasizes the need for regular pre-operation checks and predictive maintenance. Common issues associated with bearings, such as dynamic loads, poor lubrication, and misalignment, make them vulnerable under prolonged use. Despite this, the bearings also demonstrated a high MTBF, suggesting that when properly aligned, lubricated, and maintained, they can deliver stable performance with minimal recurrent failure. This dual behavior highlights the role of proper installation and routine inspection in optimizing bearing lifespan. Implementing condition-based monitoring, such as vibration sensors or thermal tracking, would further increase reliability and reduce downtime.

Table 3. Reliability metrics for component failures during Test Motion A

Component	MTTF	MTTR	MTBF
Bearing	2:15:00	0:15:00	68:08:00
Welded Joint	25:16:00	1:00:00	68:08:00
Chain	25:16:00	0:40:00	68:08:00
Sensor	66:45:00	0:30:00	68:08:00

In comparison, welded joints exhibited a significantly higher MTTF of over 25 hours, indicating greater durability. However, their MTTR was the highest (approximately 60 minutes), primarily because of the complexity of structural repair work. Fixing a welded joint typically requires specialized tools, skilled labor, and additional time for inspection and

https://doi.org/10.31436/iiumej.v27i2.3971

verification of integrity. The primary root cause of welded joint failure is fatigue accumulation under repeated mechanical loading. Because welded joints are structurally critical elements, even minor cracks can progress rapidly into major failures. Adopting non-destructive testing (NDT) techniques such as ultrasonic testing or dye-penetrant inspection during scheduled maintenance could help detect cracks at early stages. Preventing such failures would greatly reduce repair time, improve machine uptime, and eliminate the need for field welding, which is often impractical during live highway operations.

Chains, although secondary in the failure sequence, were relatively easy to repair (MTTR of 40 minutes) due to their modular nature and straightforward reconnection to gear assemblies. However, they remain vulnerable to tensile overloads and misalignment, especially if chain tension is improperly maintained. Introducing routine chain tension checks, particularly during welded joint inspections, can prevent cascading failures and preserve overall mechanical integrity. The use of automatic chain tensioners or upgraded wear-resistant materials, such as heat-treated steel, would further enhance operational resilience.

Among all components, the sensor exhibited the highest MTTF (approximately 66.75 hours), demonstrating strong reliability within the controlled laboratory environment. The sensor benefits from minimal mechanical loading and a protective casing. Nevertheless, sensors remain prone to electrical noise, vibration sensitivity, and environmental exposure, especially in real-world highway conditions. The introduction of shielded wiring, vibration-isolating mounts, and weather-resistant enclosures would enhance long-term reliability and reduce vulnerability during field operations. Although the sensor's MTTR was the shortest at 30 minutes, any malfunction still causes an immediate operational stop because the system relies heavily on accurate sensor signals to maintain continuous movement.

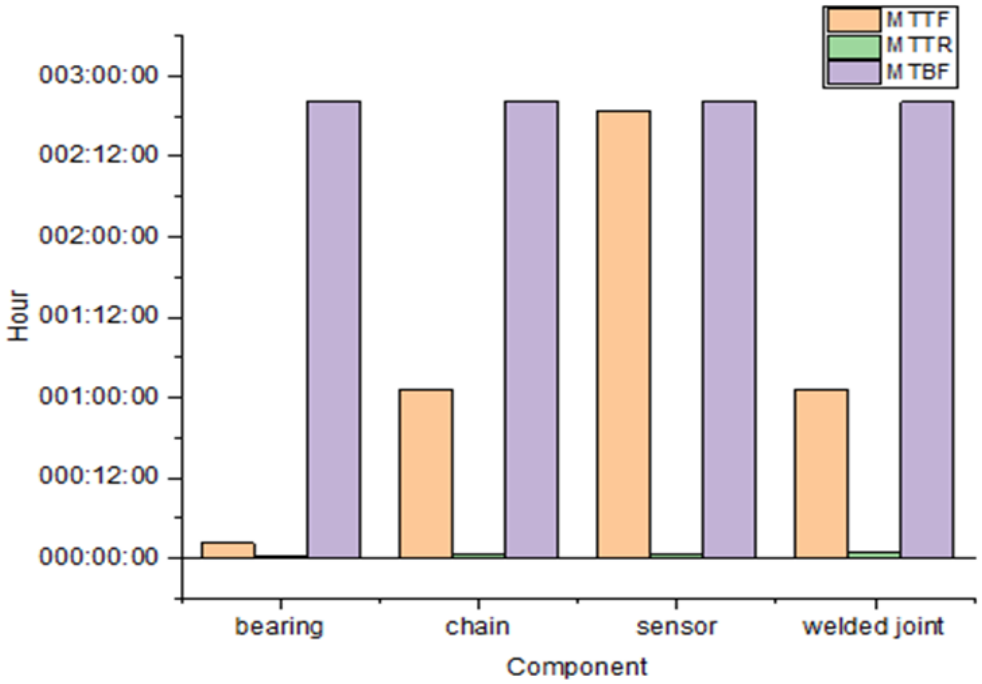


Figure 4. Comparison of MTTF, MTTR, and MTBF for C2L machine components during Test Motion A.

Figure 4 visually compares the MTTF, MTTR, and MTBF for the bearing, welded joint, chain, and sensor. It highlights the sharp contrast between the bearing's low MTTF and the much higher values for the other components. The figure also emphasizes the significantly

longer MTTR of welded joints compared to the relatively quick repair times of chains and sensors. The MTBF across all components illustrates the equal test duration rather than equivalent reliability performance. Overall, the figure effectively demonstrates component-level strengths and weaknesses, guiding engineers toward targeted areas for design enhancements and maintenance planning.

The reliability assessment of Test Motion A clearly identifies bearings and welded joints as priority components for improvement, given their susceptibility to early failure and the longer repair times required. Chains and sensors show promising reliability but still require targeted enhancements to sustain long-term field operation. Together, these findings emphasize the need for predictive maintenance, improved material selection, and expanded field data collection to refine MTBF estimates and support dependable deployment of the C2L machine in highway maintenance operations.

3.3. Component failure analysis from Test Motion B

Empirical data derived from 3,192 hours of simulated bumpy road operation during Test Motion B, as detailed in Table 4, expose critical reliability issues encountered by the C2L machine under prolonged mechanical stress. The PLC had the most breakdowns, with 249. Repairs for these failures took a total of 516 minutes. This makes the PLC the biggest problem component in the system. The PLC failure in the C2L machine stems from a hard rubber passive component used for vibration damping, insulation, or shock protection of sensitive electronics. However, this rubber degrades over time, losing flexibility and failing to absorb vibrations effectively. In dynamic environments like the C2L's cone-handling operations, continuous vibrations and impacts from moving parts are transmitted directly to the internal board, causing loose connections and cracked solder joints. These issues result in intermittent signals, logic errors, and complete PLC malfunctions during highway maintenance tasks.

Table 4. Summary of failure and repair data for C2L Machine components during Test Motion B

Component	Total failures (count)	Total Repairs (count)	Repair Time (minutes)	Total testing time (hours)
PLC	249	249	516	3192
Hook spring	1	1	110	3192

Therefore, to improve the PLC's durability, several practical solutions can be implemented. First, using vibration-isolating mounts or shock-absorbing housings such as elastomer or rubber-gasket enclosures can help reduce the amount of vibration reaching the PLC. Second, applying protective conformal coatings, such as acrylic or silicone, to the circuit boards can prevent damage from moisture and dust. Third, updating the firmware with watchdog timers and better error-handling features can help the system recover automatically from temporary faults. Together, these improvements would increase the PLC’s MTTF and make the control system more stable during real operation. On the other hand, the hook spring showed very high reliability, failing only once over the entire test period, achieving an MTTF of 3,192 hours. However, when it did fail, the repair took 110 minutes, indicating a high MTTR. Even though the failure was rare, it remains a major risk because the hook spring is integral to the cone-handling mechanism. If it breaks, the C2L machine cannot lay or collect cones, which may stop maintenance operations during busy traffic periods.

The analysis shows that the hook spring failed because the metal became tired over time, especially at the points where it was attached and repeatedly stretched and compressed. This

https://doi.org/10.31436/iiumej.v27i2.3971

fatigue worsened due to design issues such as sharp edges, uneven force distribution, and the use of materials that were not strong enough for long-term, repeated loading. To prevent this type of failure, two main improvements are recommended. First, using stronger spring materials, such as silicon-chrome, would help the spring last much longer. Second, redesigning the hook mechanism to use two springs in tandem would ensure the system continues to function even if one spring weakens. This follows TRIZ principles by adding redundancy to improve system reliability. This is approved by Becker, who highlights TRIZ as an effective framework for solving complex engineering problems and improving project outcomes [27]. Making the hook spring assembly modular with quick-release connectors would also reduce repair time and make replacement easier in the field. Overall, the failure analysis from Test Motion B shows key reliability weaknesses: the PLC is easily affected by vibration and electrical issues, and the hook spring, although rarely failing, can cause major disruption when it does. Fixing these weaknesses with proper engineering improvements will greatly increase the reliability and continuous performance of the C2L machine in real highway conditions.

3.4. Reliability metrics for failed components in Test Motion B

Table 5 presents the reliability metrics for component failures observed during Test Motion B of the C2L machine. The PLC exhibited an MTTF of 10 minutes and 12 seconds, indicating that this electronic control unit experienced relatively frequent faults during the testing period. However, its MTTR was greatly low at just 2 minutes, suggesting that the issues encountered were minor and could be swiftly rectified, likely through simple interventions such as system resets or minor adjustments. The corresponding MTBF for the PLC was recorded at 12 minutes and 16 seconds, representing the total operational interval, including both uptime and repair duration. In contrast, the hook spring exhibited markedly different reliability characteristics, with significant implications for the C2L machine's operational performance. The component recorded an MTTF of 104 hours, demonstrating exceptional durability and the ability to operate for extended periods without failure. This high reliability makes the hook spring well-suited for continuous deployment in highway maintenance operations, where sustained mechanical performance is critical. However, its MTTR was recorded at 110 minutes, reflecting the more complex, time-consuming nature of mechanical repairs when failure occurs. These metrics underscore the need for tailored maintenance strategies, frequent but quickly resolved issues affecting the PLC, and infrequent but high-impact failures linked to the Hook Spring. Both cases highlight the importance of improving component resilience and maintainability to ensure the C2L machine can meet the demanding requirements of real operational environments.

Table 5: Reliability metrics for component failures during Test Motion B

Component	MTTF	MTTR	MTBF
PLC	0:10:12	0:02:00	0:12:16
Hook spring	104:00:00	1:50:00	53:12:00

However, when failures do occur, the hook spring MTTR requires significantly more time at 1 hour and 50 minutes to complete repairs, likely involving mechanical adjustments, component replacement, or detailed fault diagnostics. The MTBF of 53 hours and 12 minutes reflects the component's overall operational reliability. These metrics indicate that although the Hook Spring is less likely to fail, its complex repair process could lead to extended machine downtime. This has the potential to disrupt highway maintenance schedules and underscores the importance of strategic maintenance planning to minimize operational interruptions during critical cone laying and collection tasks. Figure 5 illustrates the reliability metrics for two C2L

https://doi.org/10.31436/iiumej.v27i2.3971

machine components: the PLC and the hook spring. While MTTR values are not clearly visible in this figure, they represent the time required to repair a failure, thereby directly affecting equipment downtime and overall operational efficiency.

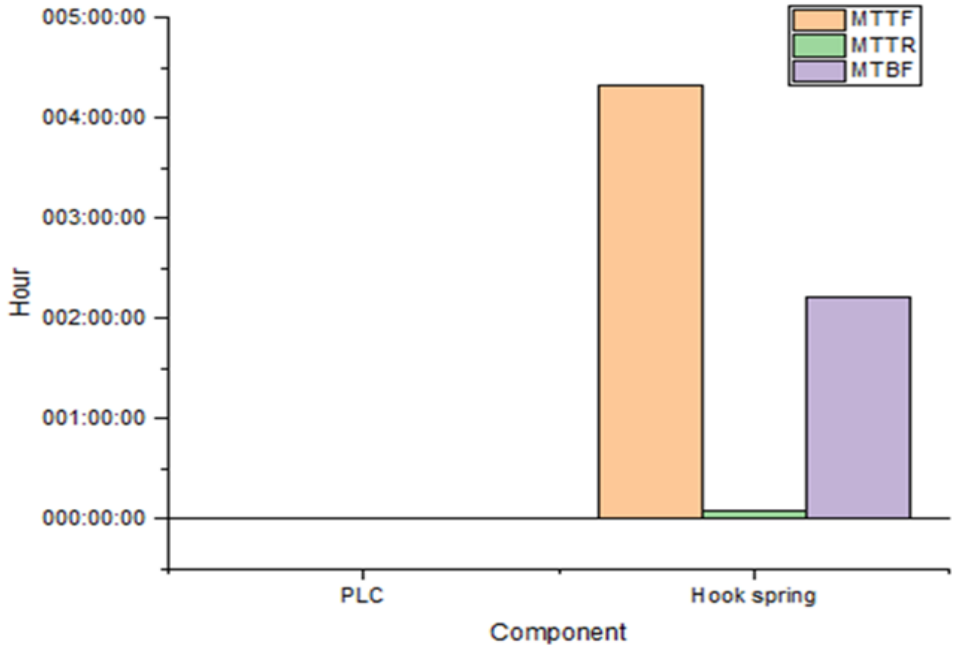


Figure 5. Comparison of MTTF, MTTR, and MTBF for C2L machine components during Test Motion B.

The reliability metrics depicted in this graph have critical implications for the operational performance of the C2L machine in highway maintenance settings. The very short MTTF of the hook spring indicates that it is prone to more frequent failures, necessitating regular maintenance. This can disrupt the machine's cone laying and collection cycles, reduce uptime, and compromise its availability during time-sensitive safety operations. The underlying cause is likely material fatigue from repetitive mechanical loading, suggesting the need for design modifications, such as stress-distributing geometries or high-tensile spring alloys.

The PLC exhibits a comparatively higher MTTF, signifying greater operational stability. However, the PLC serves as the central hub for automation and control; thus, its failure, although less frequent, results in a complete operational halt. This elevates the PLC's criticality in the system hierarchy, reinforcing the need for redundancy planning, fault-tolerant firmware, and shock-resistant housing to mitigate vibrational damage during field deployment. Then the MTBF, which integrates both the time to failure and the time to repair, offers a holistic view of each component's overall reliability in live operational conditions. Components with higher MTBF values contribute to prolonged maintenance intervals and reduced system downtime, thereby lowering lifecycle costs and enhancing deployment efficiency. Conversely, components with low MTBF values demand closer monitoring and may benefit from predictive maintenance technologies, such as sensor-based diagnostics and condition-based alerts. This integrated graph presents the reliability data for six essential components of the C2L machine, combining results from both Test Motion A and Test Motion B. The variations in MTTF, MTTR, and MTBF highlight the distinct stress responses of each component, offering crucial insights into where structural reinforcements, material upgrades, or redesigns are warranted. Understanding and acting on these reliability metrics is fundamental not only to improving operational efficiency but also to ensuring safety, minimizing unplanned downtime, and maintaining smooth traffic flow during roadside interventions.

https://doi.org/10.31436/iiumej.v27i2.3971

3.5. Evaluation and interpretation of reliability metrics for C2L machine components in Test Motion A and B

Figure 6 provides a comprehensive visualization of the reliability metrics for six critical components of the C2L machine, compiled from the combined results of Test Motion A and B. This detailed analysis utilizes MTTF, MTTR, and MTBF to assess component performance, highlighting distinct variations that critically influence the machine's overall operational efficiency. The bearing component, despite its very low MTTF of approximately 2 minutes and 15 seconds, exhibits an equally brief MTTR of around 15 seconds, resulting in a high MTBF close to 68 hours. This indicates that while bearing failures may occur relatively frequently, they are likely to wear from continuous rotational and load stresses, which can be addressed very swiftly through rapid maintenance or replacement, minimizing operational downtime. The rapid repair time reflects both the bearing's accessibility and the simplicity of its maintenance procedures.

Similarly, the chain component demonstrates a moderate MTTF of about 25 minutes and 10 seconds and a low MTTR of 40 seconds, achieving a comparable MTBF of about 68 hours. This suggests reliable mechanical power transmission, with failures such as link wear or tension loss occurring periodically but quickly remedied, ensuring consistent functionality in cone-laying operations. The sensor component presents a very high MTTF of approximately 66 hours and 45 minutes, reflecting robust operational durability under variable conditions. The short MTTR of 30 seconds further indicates that sensor-related faults are typically minor, such as signal interruptions or calibration errors, and can be resolved promptly, maintaining automation accuracy with minimal service interruption. The welded joint exhibits an MTTF similar to the chain at around 25 minutes and 10 seconds, but with a minimal MTTR, yielding an MTBF close to 68 hours. This suggests that although fatigue-related micro-cracking may arise under cyclic loading, repairs can be executed efficiently, especially when early detection enables timely intervention.

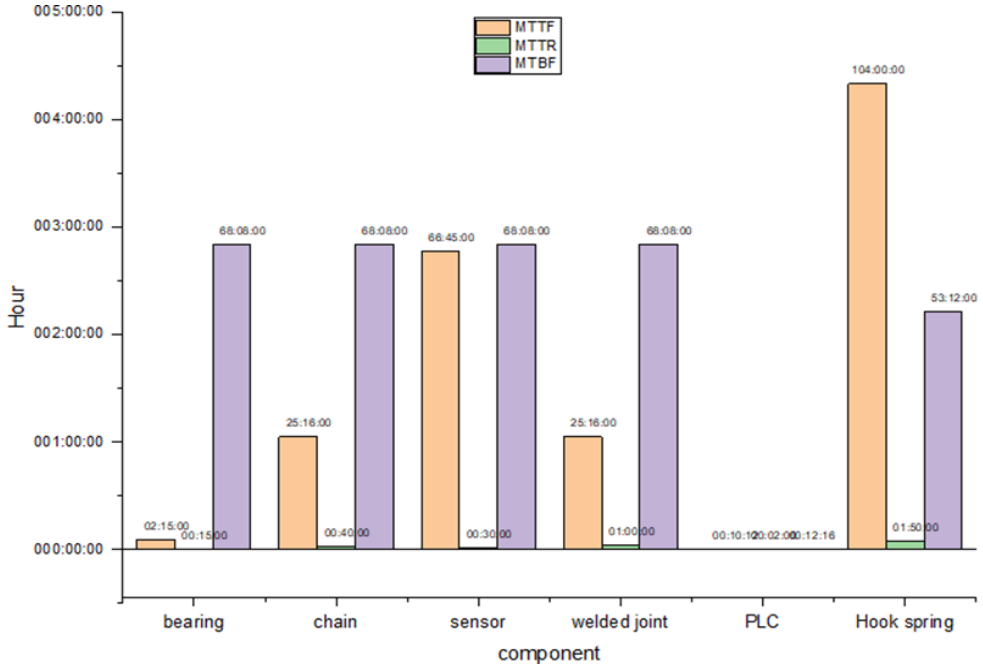


Figure 6. Summary of reliability metrics of failed components during Test Motion A and B.

The PLC component experiences frequent failures, indicated by a very low MTTF. However, its correspondingly low MTTR signifies that these electronic faults, often triggered

https://doi.org/10.31436/iiumej.v27i2.3971

by vibration-induced glitches or transient software errors, are swiftly recovered, typically through resets or minor adjustments. This balance results in an MTBF around 68 hours, reflecting a system that, while sensitive, is manageable within operational protocols. The hook spring stands out with the highest MTTF of 104 hours, underscoring its superior durability and resistance to fatigue under repeated cone-handling loads. However, its repair time is substantially longer, approximately 1 hour and 50 minutes, because of the complexity of mechanical adjustments or replacements required, often involving disassembly or precise alignment. Despite this, the hook spring maintains a robust MTBF of about 53 hours and 12 minutes, making it one of the machine's most reliable components overall. This highlights a critical trade-off between longevity and repair complexity, emphasizing the importance of predictive maintenance and potential design optimizations such as modular spring units or improved material selection to further enhance maintainability without compromising durability. Overall, these metrics provide invaluable information on component-specific vulnerabilities and maintenance needs, guiding targeted reliability improvements that can significantly boost the C2L machine's operational uptime, safety, and cost-effectiveness during highway maintenance activities.

4. CONCLUSIONS

The study clearly demonstrates that reliability metrics such as MTTF, MTTR, and MTBF are important tools for assessing the operational efficiency and maintainability of the C2L machine. Through laboratory and field testing, marked variations were observed among components in terms of failure frequency, repair times, and overall resilience under operational conditions. The sensor emerged as the most reliable component, exhibiting a high MTTF alongside a short MTTR, while the bearing, despite having the shortest MTTF, benefited from swift and straightforward repairs. In contrast, the PLC posed the greatest reliability challenges, with frequent failures and a particularly low MTTF, although its issues were often resolved through rapid resets or minor repairs. These findings represent a considerable ongoing operational cost despite its quick repair time of only 2 minutes per failure, as the cumulative downtime and repeated maintenance interventions add up to 516 minutes of total repair time, resulting in lost productivity and increased labor costs for frequent technician callouts. In contrast, the hook spring, while failing only once during the entire test period, poses a different financial challenge due to its lengthy repair time of 110 minutes, which translates to nearly two hours of machine downtime that could disrupt scheduled highway maintenance operations and potentially require expensive emergency repairs or rush-order replacement parts. Therefore, the reliability of the C2L machine depends on how well each component performs.

Based on the reliability metrics obtained from Test Motion B, the C2L machine's performance can be benchmarked against similar automated highway maintenance systems and industrial machinery operating under comparable conditions. The hook spring's MTTF of 104 hours demonstrates exceptional durability, exceeding that of typical mechanical components in mobile road equipment, suggesting it meets or surpasses industry standards for field-deployed machinery subjected to continuous mechanical loading. To improve the machine, two things must be done simultaneously: extend component life (increase MTTF) and reduce repair time and effort (reduce MTTR). Implementing proactive preventive maintenance programs, advancing component robustness through improved materials and design, and ensuring thorough training for maintenance personnel will collectively help maximize machine uptime and operational safety during critical highway maintenance tasks. Furthermore, integrating innovative methodologies, such as TRIZ-based design improvements [28], together with extended and varied testing regimes, will enable more accurate MTBF estimates and foster the

development of a more resilient, efficient, and dependable C2L machine for real-highway applications.

ACKNOWLEDGMENT

This research was conducted in collaboration between Universiti Teknikal Malaysia Melaka (UTeM) and PLUS Malaysia Berhad, and was partially funded by PLUS Malaysia Berhad, as stated in the RFA document PLUS/RFA/2019-00241/COE/AD-COR.

REFERENCES

[1] Singh Jolly S, Jit Singh B. (2014) An approach to enhance availability of repairable systems: A case study of SPMs. International Journal of Quality & Reliability Management, 31(9), 1031–1051. DOI: https://doi.org/10.1108/IJQRM-02-2014-0016.

[2] Ab-Samat H, Kamaruddin S, Abdul Aziz I. (2012) Integration of overall equipment effectiveness (OEE) and reliability method for measuring machine effectiveness. International Journal of Engineering Trends and Technology, 28(3), 150-156.

[3] Deka D, Nath T. (2015) Breakdown and reliability analysis in a process industry. International Journal of Engineering Trends and Technology, 28(3), 150–156. https://doi.org/10.14445/22315381/IJETT-V28P229.

[4] Saritha G, Tirumala Devi M., Uma Maheswari TS. (2020) Reliability and availability for non-repairable & repairable systems using Markov modelling. International Journal of Engineering Research & Technology (IJERT), Volume 09, Issue 03.

[5] Mehdi I, Boudi EM, Mehdi MA. (2024) Reliability, availability and maintainability assessment of a mechatronic system based on timed colored Petri nets. Applied Sciences, 14(11), 4852. DOI: 10.3390/app14114852.

[6] O'Connor PD, Kleyner A. (2012) Practical reliability engineering (5th ed.). John Wiley & Sons.

[7] Blanchard BS, Fabrycky WJ. (2011) Systems engineering and analysis (5th ed.). Prentice Hall.

[8] Nurdiansyah D, Hariadi ADP. (2023) Analysis of maintenance management at PT. XYZ power plant (with MTTR, MTTF, and availability factors) and development of performance improvement program. FJMR (Fakultas Jurnal Manajemen dan Riset), 3(9), https://doi.org/10.55927/fjmr.v3i9.11137.

[9] Ebeling CE. (2019) An introduction to reliability and maintainability engineering (3rd ed.). Waveland Press.

[10] Dhillon BS. (2006) Maintainability, maintenance, and reliability for engineers. CRC Press.

[11] Smith DJ, Hinchcliffe GR. (2004) Routledge handbook of maintenance management and engineering. Routledge.

[12] Ben-Daya M, Duffuaa SO, Raouf A, Knezevic J, Ait-Kadi D (Eds.). (2009) Handbook of maintenance management and engineering. Springer, 2009. DOI: 10.1007/978-1-84882-472-0.

[13] Facilio. (2014) Understanding MTBF and its role in maintenance management. Facilio Insights. Retrieved from https://www.facilio.com/insights/mtbf-maintenance-management.

[14] Salawu EY, Awoyemi O, Akerekan OE, Afolalu AS. (2023) Impact of maintenance on machine reliability: A review. E3S Web of Conferences, 430, Article 01226, 2023. DOI: 10.1051/e3sconf/202343001226.

[15] Parida A, Kumar U. (2006) Maintenance performance measurement (MPM): Issues and challenges. Journal of Quality in Maintenance Engineering, 12(3), 239–251. https://doi.org/10.1108/13552510610685084.

[16] Campbell JD, Reyes-Picknell JV. (2015) Uptime: Strategies for excellence in maintenance management" (3rd ed.). Productivity Press.

[17] International Electrotechnical Commission. (2009) IEC 60300-3-11: Dependability management – Part 3-11: Application guide"- Reliability centered maintenance.

[18] Ebeling CE. (2010) An introduction to reliability and maintainability engineering (2nd ed.). Waveland Press.

[19] Ghodrati B. (2005) Reliability and operating environment-based spare parts planning. Luleå University of Technology.

[20] Narang RV. (2012) Introducing reliability and maintainability in engineering and technology. American Society for Engineering. Education. https://peer.asee.org/introducing-reliability-and-maintainability-in-engineering-and-technology.pdf.

[21] Ben-Daya M. (2022) Advances in maintenance optimization. Springer.

[22] Pricop E, Mihalache SF, Paraschiv N. (2016) Considerations regarding security issues impact on systems availability. Electronics, Computers and Artificial Intelligence Conference, 30 June–02 July, Ploiesti, Romania.

[23] Atlassian. (2024) Reliability vs. availability: Key metrics for system performance. Available https://www.atlassian.com/incident-management/kpis/reliability-vs-availability.

[24] Lacarella JC, Chea R, Patterson DA, Weller JD. (2024) Projecting exceedance of juvenile salmonid thermal maxima in streams under climate change: A crosswalk from lab experiments to streamside restoration. Freshwater Biology 69, no. 9: 1218–1231.

[25] Żyluk A, Zieja M, Grzesik N, Tomaszewska J, Kozłowski G, Jasztal M. (2023) Implementation of the mean time to failure indicator in the control of the logistical support of the operation process. Applied Sciences, 13(7), 4608. https://doi.org/10.3390/app13074608.

[26] Yang X, Zhang J, Liu W, Jing J, Zheng H, Xu W. (2024) Automation in road distress detection, diagnosis and treatment. Journal of Road Engineering 4, no. 1 (2024): 1–26. https://doi.org/10.1016/j.jreng.2024.01.005.

[27] Becker F. (2024). Engineering problems solving method (TRIZ) for TIC projects. Master's thesis, Universitat Politècnica de Catalunya.

[28] Ismail A, Salim MA, Mohd Yusof SW, Chew KW. (2023). Optimization of cone lifter componentsin C2L machine for highway concession operation using TRIZ method. International Journal of Integrated Engineering 16, no. 9: 182–189. https://penerbit.uthm.edu.my/ojs/index.php/ijie/article/view/17721.

https://doi.org/10.31436/iiumej.v27i2.4013

Analytical and Flight Test-Based Dynamic Characteristic Identification of a Small Fixed-Wing UAV

ERIES BAGITA JAYANTI[1*], YAZDI IBRAHIM JENIE[2], ARDIAN RIZALDI[1,3], NOVITA ATMASARI[1], FUAD SURASTYO PRANOTO[1]

[1]*Research Center for Aeronautics Technology, National Research and Innovation Agency, Bogor, Indonesia*
[2]*Aerospace Engineering, Faculty of Mechanical and Aerospace Engineering, Institut Teknologi Bandung, Bandung, Indonesia*
[3]*Aerospace Engineering, Graduate School of Mechanical and Aerospace Engineering, Gyeongsang National University, Jinju, Gyeongsangnam-do, Korea*

**Corresponding author: erie002@brin.go.id*

(Received: 27 October 2025; Accepted: 9 March 2026; Published online: 10 May 2026)

ABSTRACT: Reliable and efficient UAV flight control design requires accurate knowledge of the vehicle's dynamic characteristics, which are rarely provided by manufacturers and must therefore be determined independently. These characteristics can be obtained through analytical calculations, wind tunnel testing, and flight testing. However, in small UAV development, the process is often simplified by skipping directly to flight tests for control tuning, without first examining the underlying dynamic behavior. To address this gap, this research compares the dynamic characteristics derived from analytical calculations and flight testing. The case study focuses on BRIN's bMFE-UAV, in which longitudinal and lateral/directional dynamics are identified using control-surface deflections as inputs and measured responses of speed, altitude, attitude, acceleration, and rotational rate as outputs. Results reveal consistent oscillatory behaviors: two oscillatory modes in longitudinal dynamics and, in lateral/directional dynamics, one oscillatory mode accompanied by two non-oscillatory modes. Only the damping ratio of the short-period mode showed close agreement (<10% difference), while other parameters exhibited substantial discrepancies. The discrepancies demonstrate the limitations of analytical methods in accurately reflecting real flight dynamics, reinforcing the necessity of flight testing to ensure reliable data for small UAV flight-control design.

ABSTRAK: Reka bentuk kawalan penerbangan UAV yang boleh dipercayai dan cekap memerlukan pengetahuan tepat pada ciri dinamik kenderaan. Ianya jarang didapati daripada pengeluar dan oleh itu mesti ditentukan secara bebas. Ciri-ciri ini boleh diperolehi melalui pengiraan analitika, ujian terowong angin, dan ujian penerbangan. Walau bagaimanapun, dalam pembangunan UAV kecil, prosesnya sering dipermudahkan dengan melangkau terus ke ujian penerbangan bagi kawalan penalaan, tanpa terlebih dahulu memeriksa tingkah laku dinamik. Kajian ini bertujuan membandingkan ciri dinamik yang diperoleh daripada pengiraan analitika dan ujian penerbangan. Kajian kes memfokuskan pada bMFE-UAV BRIN, di mana dinamik membujur dan sisi/arah dikenal pasti menggunakan pesongan permukaan kawalan sebagai input dan tindak balas diukur pada kelajuan, ketinggian, sikap, pecutan dan kadar putaran sebagai hasil pengeluaran. Dapatan mendapati tingkah laku berayun yang konsisten: dua mod berayun dalam dinamik membujur dan dalam dinamik sisi/arah, satu mod berayun disertai dengan dua mod bukan berayun. Hanya nisbah redaman mod jangka pendek menunjukkan perbezaan terdekat (iaitu <10%), manakala parameter lain menunjukkan ralat yang besar. Ralat menunjukkan batasan kaedah analisis dalam

mencerminkan dinamik penerbangan sebenar, mengukuhkan keperluan ujian penerbangan bagi memastikan data boleh dipercayai untuk reka bentuk kawalan penerbangan UAV kecil.

KEYWORDS: *Dynamic Characteristic Identification, Dynamic Modeling, Flight Test, UAV*

1. INTRODUCTION

Accurate knowledge of UAV dynamic characteristics is essential for the design of reliable and efficient flight control systems. However, UAV manufacturers rarely provide such data, necessitating independent investigations. The three most common methods for gathering aircraft dynamic characteristics are analytical calculations, wind-tunnel testing, and flight testing. The analytical calculation, using available software such as the USAF DATCOM or Computational Fluid Dynamics (CFD), involves many assumptions and simplifications, leading to inaccurate results [1, 2]. However, this method is the quickest, safest, and cheapest of the three and is ideal for preliminary research [3, 4]. Utilizing a wind tunnel and a scaled model enables more accurate aerodynamic data to be obtained. It can be used to derive dynamic characteristics data [5, 6]. However, Wind Tunnel testing is expensive and time-consuming, not to mention the rarity of the facilities. The third method is to conduct flight tests directly, gathering the aircraft's dynamic characteristics in the actual flight environment. While this method is considered the most accurate, it requires considerably more resources and time, and poses greater risks to the aircraft, pilots, and the environment [7, 8].

In crewed aircraft development, these methods are commonly used in succession, from the quickest and cheapest to the most time-consuming and expensive. The resulting data is updated continuously once results from more accurate methods are obtained. However, the paradigm changes in small UAV development. The development costs and operational risks are much lower. Typically, small UAV developers skip complex analysis and wind-tunnel testing, proceeding directly to flight testing and iterating on the vehicle control design based on experience [9]. Many failures that destroy prototypes are common during the iteration, yet still cheaper than conducting a rigorous derivation, as in crewed aircraft development. Nevertheless, the process skips many engineering steps, including deriving dynamic characteristics. The lack of documentation, resulting from skipped steps, also invalidates the control system design and hinders further development [10]. This condition somewhat contradicts the role of flight testing, which is generally regarded as the most accurate method for deriving dynamic characteristics. These characteristics are typically used as the basis for flight control system design. Therefore, a direct comparison between analytically derived dynamic characteristics and those obtained from flight-test-based identification is necessary to assess the reliability and applicability of analytical models for small UAVs.

Based on this motivation, the objective of this research is to perform a comparative analysis of UAV dynamic characteristics obtained from analytical calculations and flight test-based identification. Such a methodology has been demonstrated frequently in many studies [11-13], but is rarely applied in small-UAV development, which often relies on trial and error to determine appropriate control gains. The case study in this research is the MFE surveillance UAV owned by the National Research and Innovation Agency (BRIN), hereafter referred to as the bMFE-UAV. Its dynamic characteristics are identified along both the longitudinal and lateral/directional axes using recorded flight-test data [13-15]. The deflections of the UAV control surfaces, including its elevator and ailerons, are used as input variables, while the recorded IMU and other sensor data, including vehicle speed, altitude, attitude, acceleration, and rotational rate, serve as the output for the identification process. The results are then compared with analytical data obtained using the USAF DATCOM software.

https://doi.org/10.31436/iiumej.v27i2.4013

This article is structured as follows. The case study UAV will be explained along with its specifications. Afterward, the vehicle dynamic characteristics are derived by first determining its mathematical model, and aerodynamic data is obtained using the USAF DATCOM. Then, it describes the vehicle's flight test process, the essential data obtained, and the identification process that used Ordinary Least Squares analysis. All the results are then compared, and the research concludes with several suggestions.

2. METHODOLOGY

2.1. Case Study: bMFE-UAV

Figure 1 shows the bMFE-UAV used in this study, a twin-engine fixed-wing aircraft with a conventional configuration.



Figure 1. bMFE-UAV.

The UAV is powered by two electric engines that drive two propellers on its wings. The motion of rolling, pitching, and yawing is actuated by its ailerons on its wing and elevators and rudder on its T-Tail configured stabilizer. The UAV is mainly made of Styrofoam, with polymer reinforcement at the joints and in its engine compartment. It is widely used in topographical mapping, land surveying, precision farming, environmental monitoring, and many other missions. Table 1 lists further specifications of the bMFE-UAV [16].

Table 1. Specification of UAV

Specifications	Symbol	Value
Wingspan	b	2.43 m
Wing Area	S	0.75 m^2
Mean Aerodynamic Chord	$\bar{c}$	0.3 m
Horizontal tail area	S_H	0.07 m^2
Vertical tail area	S_V	0.065 m^2
Max. Take-off Weight	MTOW	11.5 kg
Suggested payload	$m_{payload}$	< 2.5 kg
Cruise flight speed	V	17-20 m/s
Cruising altitude	h	300 m
Longest flight range	RNG	150 km
Moment of inertia around x axis	I_x	0.8244 kg.m^2
Moment of inertia around y axis	I_y	1.1350 kg.m^2
Moment of inertia around z axis	I_z	0.1723 kg.m^2
Moment of inertia in xz axis	I_{xz}	0.1266 kg.m^2

https://doi.org/10.31436/iiumej.v27i2.4013

2.2. Dynamics Characteristics Derivation by Analytical Calculations

This section explains the dynamic characteristics derived from analytical methods, primarily based on data computed with the USAF DATCOM software.

2.2.1. Aircraft Equation of Motion

The equation of motion of any moving body (including Aircraft) is derived from Newton's Laws of Motion for a six degrees of freedom (6-DOF) motion in the Three-Dimensional Space [17]. These are sets of translational and rotational equations of motion, accompanied by kinematic equations, represented in Eqs. (1), (2), and (3), respectively. All vectors are defined in the aircraft stability axis on cruising conditions.

$$\frac{d\bar{V}}{dt} = \frac{\sum \bar{F}}{m} - \bar{\omega} \times \bar{V} \tag{1}$$

$$\frac{d\bar{\omega}}{dt} = \boldsymbol{I}^{-1}[\sum \bar{M} - \bar{\omega} \times \boldsymbol{I}\bar{\omega}] \tag{2}$$

$$\begin{bmatrix} \dot{\varphi} \\ \dot{\theta} \\ \dot{\psi} \end{bmatrix} = \begin{bmatrix} 1 & sin\,\varphi\, tan\,\theta & cos\,\varphi\, tan\,\theta \\ 0 & cos\,\varphi & -sin\,\varphi \\ 0 & \frac{sin\,\varphi}{cos\,\theta} & \frac{cos\,\varphi}{cos\,\theta} \end{bmatrix} \bar{\omega} \tag{3}$$

where the vectors of translational and rotational rates are defined as:

$$\bar{V} = [u \quad v \quad w]^T \tag{4}$$

and,

$$\bar{\omega} = [p \quad q \quad r]^T \tag{5}$$

Mass (m) is assumed to be constant during the analysis and set to the actual flight-test weight in Table 5. The moment of inertia (I) is a tensor matrix defined as in Eq. (6), where the polar inertia is listed in Table 1, and all cross-inertias are assumed to be zero, except for I_{xz}.

$$\boldsymbol{I} = \begin{bmatrix} I_x & -I_{xy} & -I_{xz} \\ -I_{xy} & I_y & -I_{yz} \\ -I_{xz} & -I_{yz} & I_z \end{bmatrix} \tag{6}$$

The forces and moments acting on the aircraft are modeled as a summation of the aspects of aerodynamics, propulsion, and gravitational (weight), as in Eq. (7) and Eq. (8), referred to by the superscripts A, P, and W, respectively.

$$\sum \bar{F} = \bar{F}^A + \bar{F}^P + \bar{F}^W \tag{7}$$

$$\sum \bar{M} = \bar{M}^A + \bar{M}^P \tag{8}$$

The sum of moments in Eq. (8) is formulated based on the aircraft's center of gravity. As a result, weight does not contribute to the moment and is therefore excluded from the moment equation.

These forces and moments are detailed in the following paragraphs:

a. Aerodynamics

The bMFE-UAV aerodynamic forces and moments are modeled in Eq. (9) and Eq. (10). S denotes the reference or wing planform area, $\bar{c}$ and b is the mean aerodynamic chord and wingspan length, and V denotes the vehicle airspeed. The values of those parameters are listed in Table 1.

https://doi.org/10.31436/iiumej.v27i2.4013

Eq. (9) is initially formulated in the aerodynamic axis system and subsequently transformed into the body axis using the transformation matrix defined in Eq. (11). In this formulation, the aerodynamic lift and drag directions are defined opposite to the positive body-axis directions, whereas the side force follows the positive body y-axis convention.

$$\bar{F}^A = \frac{1}{2}\rho V^2 S\, C_b^a \begin{bmatrix} C_D\{\alpha\} \\ C_Y\{\beta, \delta_a\} \\ C_L\{\alpha, \delta_e\} + C_{L_q}\frac{\bar{c}}{2V}q \end{bmatrix} \tag{9}$$

$$\bar{M}^A = \frac{1}{2}\rho V^2 S \begin{bmatrix} b\left(C_\ell\{\beta, \delta_a\} + C_{\ell_p}\frac{b}{2V}p + C_{\ell_r}\frac{b}{2V}r\right) \\ \bar{c}\left(C_m\{\alpha, \delta_e\} + C_{m_q}\frac{\bar{c}}{2V}q\right) \\ b\left(C_n\{\beta, \delta_a\} + C_{n_p}\frac{b}{2V}p + C_{n_r}\frac{b}{2V}r\right) \end{bmatrix} \tag{10}$$

$$C_b^a = \begin{bmatrix} -\cos\alpha\cos\beta & -\cos\alpha\sin\beta & \sin\alpha \\ -\sin\beta & -\cos\beta & 0 \\ -\sin\alpha\cos\beta & -\sin\alpha\sin\beta & -\cos\alpha \end{bmatrix} \tag{11}$$

In this study, the aerodynamic model formulation conceptually accounts for rudder effects; however, rudder-related aerodynamic derivatives are not estimated. This is because the flight test campaign did not include deliberate rudder excitation, resulting in insufficient excitation for reliable parameter identification. Consequently, rudder contributions are not explicitly included in Eq. (9) and Eq. (10).

The symbol ρ denotes the atmospheric density, which this research will use the value 1.18 kg/m3, representing the UAV cruising altitude at 300 m. The forces in Eq. (9) undergo axis transformation into the aircraft body axis, via the transformation matrix, C_b^a, presented in Eq. (11). The aerodynamic derivatives in Eq. (9) and Eq. (10) include the effect of steady-state and perturbed flight conditions. The steady-state coefficients are the derivatives for forces of drag, side force, and lift, or C_D, C_Y, and C_L, and the derivatives for moments of rolling, pitching, and yawing, or C_ℓ, C_m, and C_n, respectively. The value for these coefficients depends on the steady-state flight conditions: the angle of attack (α) and the side slip angle (β), and the steady-state control input: the elevator deflection (δ_e) and aileron deflection (δ_a).

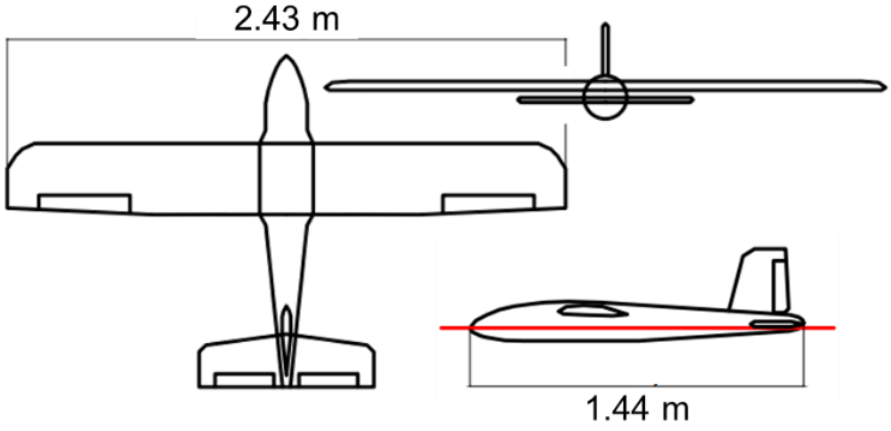


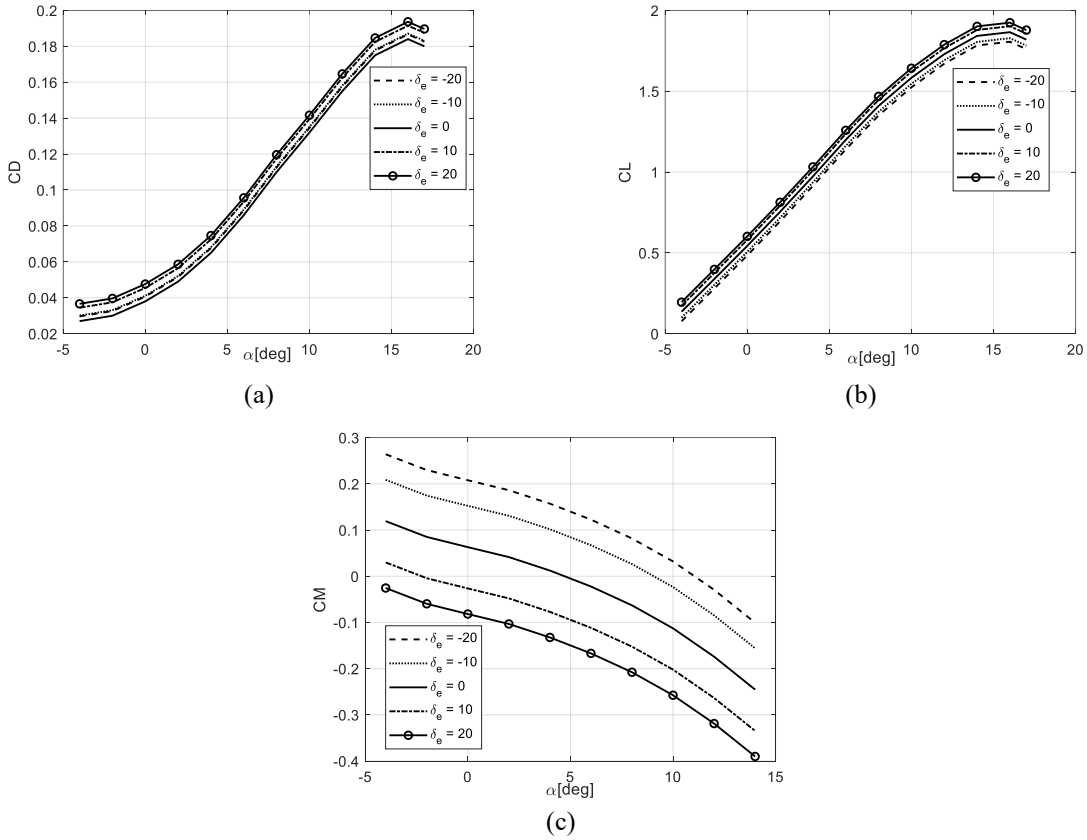
Figure 2. Three-view drawing of UAV, simplified for DATCOM.

Eq. (9) and Eq. (10) also account for aerodynamics under perturbed flight conditions, represented by the dynamic derivatives of force and moment with respect to the rotational rates

https://doi.org/10.31436/iiumej.v27i2.4013

(p, q, and r). These include the lift and pitching-moment derivatives with respect to the pitch rate, and the rolling and yawing moment derivatives with respect to the roll and yaw rates. Those aerodynamic derivatives are obtained using the USAF Digital DATCOM software, which requires the geometric description of bMFE-UAV configurations and flight conditions. The outputs will be the complete set of dimensionless aerodynamic derivatives required for Eqs. (9) and (10). The basic geometry of the bMFE-UAV is listed in Table 1, which shows the conventional configuration comprising a fuselage, wing, horizontal tailplane, and vertical tailplane. Other dimensions and configurations required for DATCOM input were obtained by direct measurement of the vehicle and from the three-view drawing in **Figure 2**.

Figure 3 (a) to (c) depict the results of longitudinal aerodynamic derivatives due to variation of the angle of attack (α), in various elevator deflection (δ_e). **Figure 4** (a) to (c), on the other hand, shows the lateral/directional due to the side slip angles (β). Only the roll coefficient (C_ℓ) is affected by the aileron deflection (δ_a).



(a) (b)

(c)

Figure 3. Aerodynamic characteristics of the bMFE-UAV, including the (a) Lift force coefficient; (b) Drag force coefficient; (c) Pitch moment coefficient.

The dynamic derivatives in longitudinal and lateral/directional modes are presented in Table 2, with the symbols used in Eqs. (9) and (10). These values are assumed to be constant within the flight-condition dynamics during the analysis. Other rotational-rate effects on the forces and moments may also occur, but are assumed to be insignificant and neglected.

https://doi.org/10.31436/iiumej.v27i2.4013

Table 2. Dynamics Derivatives Coefficient

Parameter	Symbol	Value [./rad]
Lift derivative due to the pitching rate	C_{L_q}	6.145
Pitch moment derivative due to the pitching rate	C_{m_q}	-11.236
Roll moment derivative due to the rolling rate	C_{ℓ_p}	-0.513
Roll moment derivative due to the yawing rate	C_{ℓ_r}	0.149
Yaw moment derivative due to the rolling rate	C_{n_p}	-0.044
Yaw moment derivative due to the yawing rate	C_{n_r}	-0.046

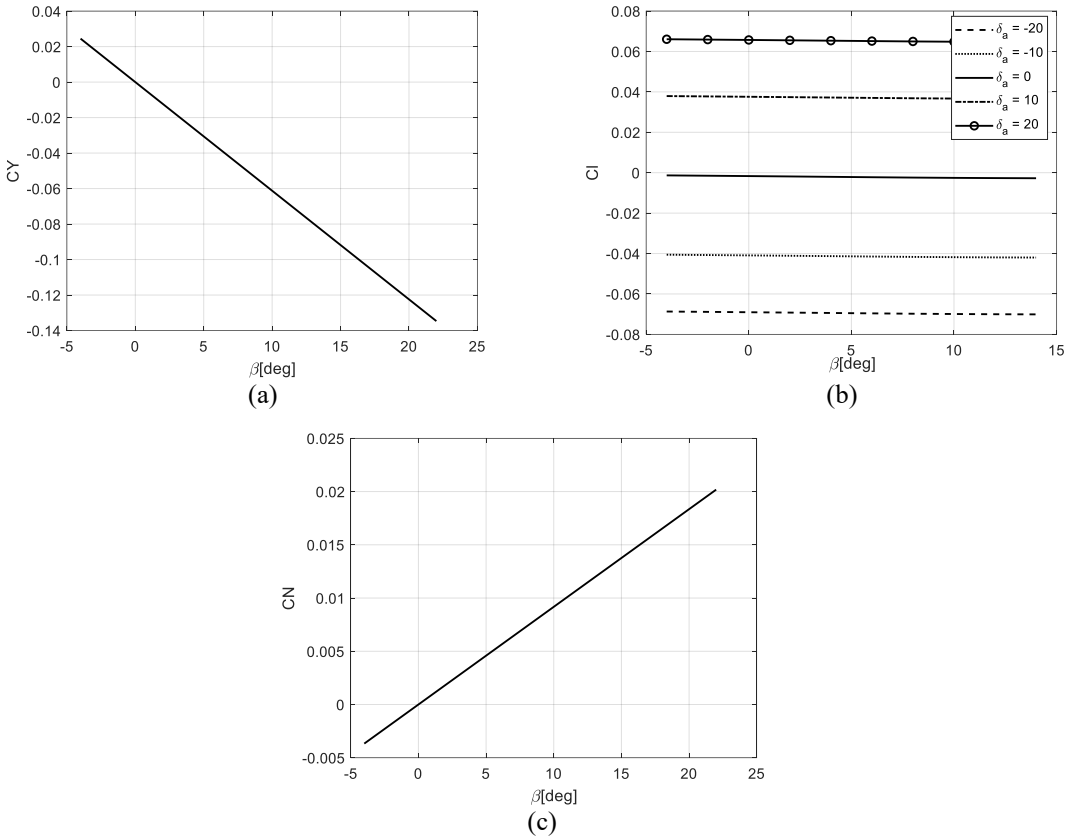


Figure 4. Aerodynamic characteristics of the bMFE-UAV, including the (a) side-force moment coefficient; (b) roll moment coefficient; and (c) yaw moment coefficient.

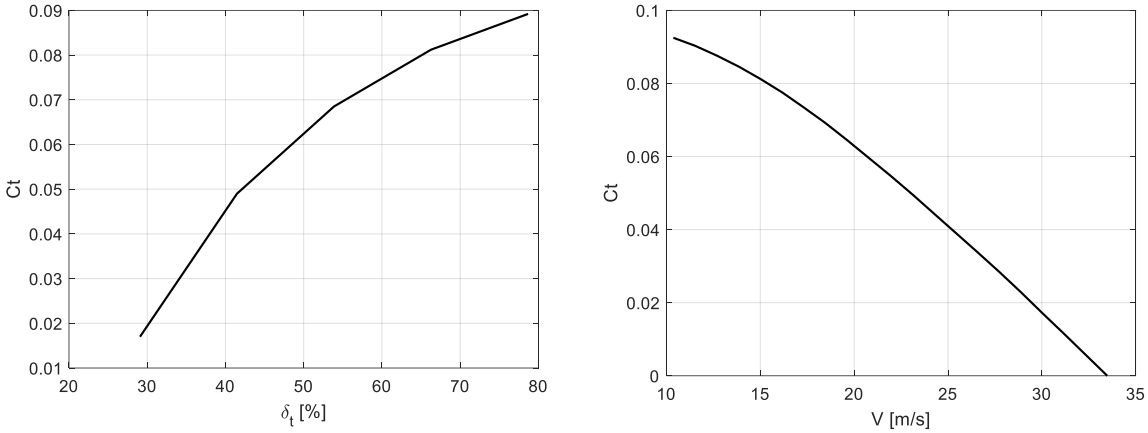
b. Propulsion

The bMFE-UAV motors are the 3520-500 KV model, with a no-load current of 1.1 A and a resistance of 30 mW. The propellers are two-bladed APC1510 series with 15-inch diameter and 10-inch pitch. The propulsion lines of bMFE, from its two electric propeller engines, have zero incidence angle from the body axis. Therefore, the propulsion only generates forces in the x-body axis, as shown in Eq. (12). The moment equation in Eq. (13), where d_T represents the distance between the thrust and the aircraft's c.g.

$$\bar{F}^P = \frac{1}{2}\rho V_s^2 S_{ref} \begin{bmatrix} C_T\{\delta_T\} + C_{T_u}\hat{u} \\ 0 \\ 0 \end{bmatrix} \tag{12}$$

https://doi.org/10.31436/iiumej.v27i2.4013

$$\bar{M}^P = \frac{1}{2}\rho V_s^2 S_{ref} d_T \begin{bmatrix} 0 \\ C_T\{\delta_T\} + C_{T_u}\hat{u} \\ 0 \end{bmatrix} \quad (13)$$



(a) (b)

Figure 5. (a) Thrust Coefficient due to Throttle Deflection; (b) Thrust Coefficient due to velocity.

The propulsion is derived further as a steady-state thrust coefficient C_T, which depends on the throttle input, δ_T, and the dynamic coefficient due to the forward speed u. These are obtained mainly from the bMFE-UAV datasheets and are then shown in **Figure 5**.

c. Gravitational Weights

The last aspect in the definition of the forces in Eq. (7) is the weight aspect, which is modeled as the constant mass (m) multiplied by the gravitational acceleration (g) and transformed to the body axis, as presented in Eq. (14).

$$\bar{F}^W = mg \begin{bmatrix} -\sin\theta \\ \sin\varphi\cos\theta \\ \cos\varphi\cos\theta \end{bmatrix} \quad (14)$$

2.2.2. Non-Linear Simulink Representation

One way to evaluate the vehicle dynamics from the data is to simulate the bMFE-UAV computationally. The equations from Eq. (1) to (14) can be represented using the Simulink Block Set, as illustrated in **Figure 6**. The calculations of the aerodynamic steady-state coefficient are done using the 2-dimensional look-up table block, based on the corresponding aerodynamic angles and control surface deflection.

The UAV dynamic response can be evaluated using this Simulink model, as shown in **Figure 7** for (a) longitudinal and (b) lateral/directional motions. To assess longitudinal dynamics, the UAV model was given a 22° elevator deflection at a steady-state velocity of 15 m/s for 1 s. On the other hand, to trigger the lateral/directional responses, the model was subjected to an 11° aileron deflection doublet for 2 seconds from an 18 m/s steady-state velocity. The values of the initial parameter and the inputs are chosen to match those used in the experiments. Both modes are simulated for 20 seconds, with inputs provided at the 5-second time mark after steady-state conditions are achieved.

Evidently, the responses match the typical dynamic response for conventional aircraft, which are stable in longitudinal and lateral/directional motions. On the longitudinal, two distinguishable stable oscillations can be observed in **Figure 7** (a), which typically are known as the short-period mode oscillation (high frequency – high damping ratio) and the phugoid

mode oscillation (low frequency - low damping ratio). On the other hand, the lateral/directional motions show only one type of oscillation, most likely the Dutch roll mode, which combines stable rolling and yawing motions. This oscillation is most likely accompanied by other stable, non-oscillating motions that are harder to observe in lateral/directional responses in **Figure 7**(b).

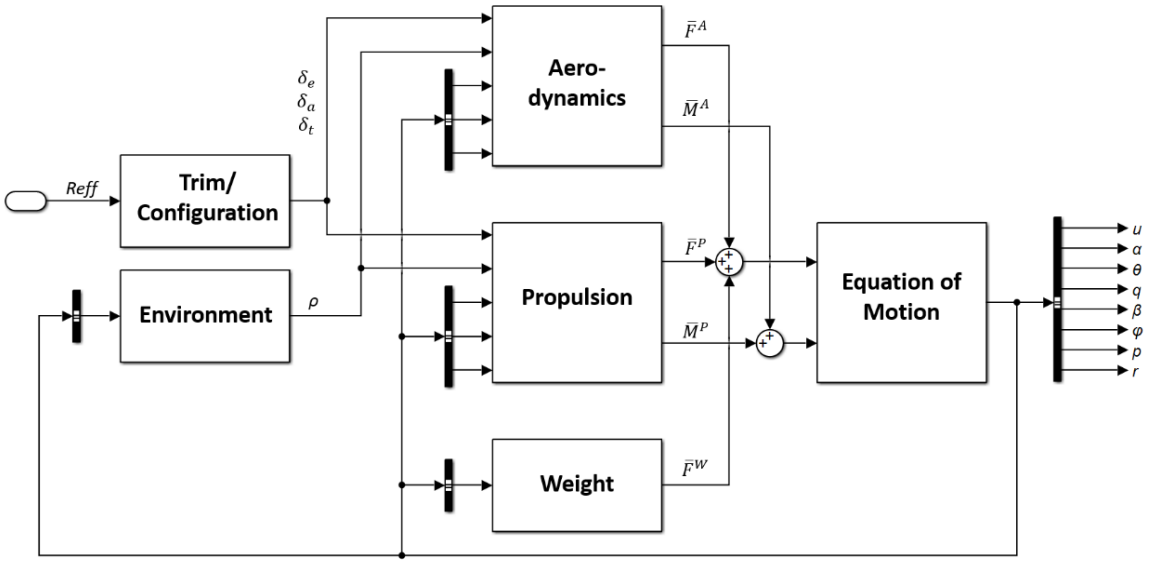
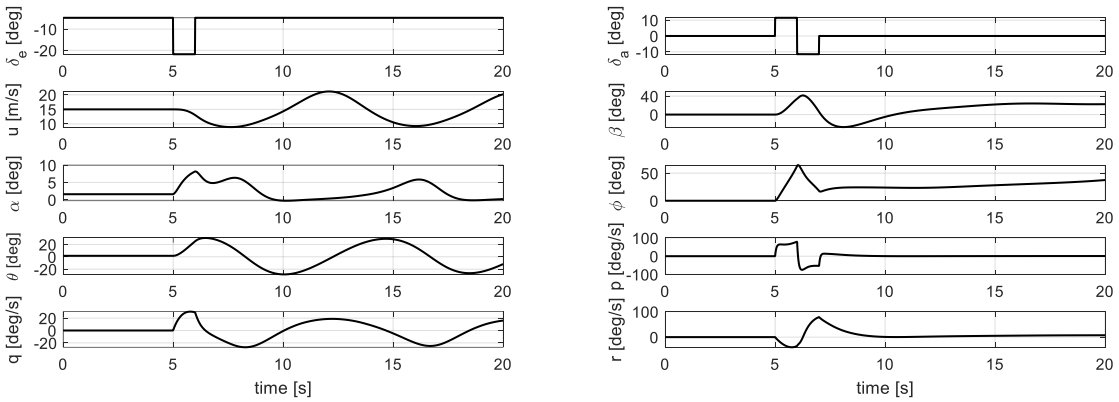


Figure 6. Simulink pseudo-model for the bMFE-UAV dynamics.



(a) (b)

Figure 7. (a) Longitudinal dynamic responses of the bMFE-UAV; (b) Lateral/directional dynamic responses of the bMFE-UAV

2.2.3. Linear State-Space Representation

Another way to evaluate the bMFE-UAV dynamic response analytically is to retrieve its dynamic parameters from the linearized equation of motion. This can be done using Simulink's Model Linearizer tools, from the previous model depicted in **Figure 6**. Adjusted for the typical states used in aircraft dynamic analysis [17], the linear state-space representation is shown in Eq. (15)-(17) for the longitudinal response, and Eq. (18)-(20) for the lateral directional response.

$$\dot{\bar{x}}_{\text{lo}} = A_{\text{lo}}\bar{x}_{\text{lo}} + B_{\text{lo}}\bar{u}_{\text{lo}},\ \bar{x}_{\text{lo}} = [u \quad w \quad \theta \quad q]^T,\ \bar{u}_{\text{lo}} = [\delta_e \quad \delta_t]^T \tag{15}$$

https://doi.org/10.31436/iiumej.v27i2.4013

$$A_{\text{lo}} = \begin{bmatrix} -0.238 & 0.5468 & -9.802 & -0.3993 \\ -1.167 & -5.283 & -0.2724 & 14.37 \\ 0 & 0 & 0 & 1 \\ 0.1114 & -0.1729 & 0 & -1.954 \end{bmatrix} \tag{16}$$

$$B_{\text{lo}} = \begin{bmatrix} -0.1387 & 0.0934 \\ -1.587 & 0.0045 \\ 0 & 0 \\ -4.607 & -0.0143 \end{bmatrix} \tag{17}$$

$$\dot{\bar{x}}_{\text{ld}} = A_{\text{ld}}\bar{x}_{\text{ld}} + B_{\text{ld}}\bar{u}_{\text{ld}}, \bar{x}_{\text{ld}} = [v \quad \varphi \quad p \quad r]^T, \bar{u}_{\text{ld}} = [\delta_a \quad \delta_r]^T \tag{18}$$

$$A_{\text{ld}} = \begin{bmatrix} -0.296 & 9.8 & -0.223 & -17.999 \\ 0 & 0 & 1 & -0.011 \\ -3.61 & 0 & -15.638 & 5.960 \\ 8.636 & 0 & 5.350 & -10.676 \end{bmatrix} \tag{19}$$

$$B_{\text{ld}} = \begin{bmatrix} 0 & 3.22 \\ 0 & 0 \\ 105.6 & 33 \\ -112.8 & -71.19 \end{bmatrix} \tag{20}$$

Table 3. Dynamic characteristics of bMFE-UAV, obtained from the linearized longitudinal equation of motions

	Parameter	Value
Characteristic Polynomial	$\Delta_{lo} = s^4 + 7.48s^3 + 15.21s^2 + 4.78s + 7.75$	
Short Period	Characteristics Roots	-3.7111 ± 0.710i
	Natural Frequency [rad/s]	3.7785
	Damping Ratio	0.9822
Phugoid	Characteristics Roots	-0.0262 ± 0.736i
	Natural Frequency [rad/s]	0.7368
	Damping Ratio	0.0356

Table 4. Dynamic Characteristics of bMFE-UAV, obtained from the linearized lateral/directional equation of motions

	Parameter	Value
Characteristic Polynomial	$\Delta_{ld} = s^4 + 26.6s^3 + 297.5s^2 + 2162.3s - 114.2$	
Dutch Roll	Characteristics Roots	-5.053 ± 10.230i
	Natural Frequency [rad/s]	11.47
	Damping Ratio	0.445
Roll Subsidence	Characteristics Roots	-16.557
	Half-life time (s)	0.0419
Spiral	Characteristics Roots	0.053
	Doubling time (s)	13.1825

From the dynamic matrices A_{lo} and A_{ld}, the characteristic polynomials of the longitudinal and lateral/directional motion can be derived, from which the dynamics parameter can be obtained. These results are then listed in Table 3 for longitudinal motions and Table 4 for the lateral/directional.

Confirming the results from the non-linear simulation in **Figure 7**, the dynamic parameters from the linearized equation of motion also show typical dynamic responses for a conventional aircraft. Two complex-conjugate pairs of characteristic roots are retrieved in the longitudinal

https://doi.org/10.31436/iiumej.v27i2.4013

direction, as shown in Table 3. These indicate two oscillatory dynamics that match the typical short-period and phugoid modes. While both are dynamically stable, the damping ratio of the phugoid mode might be inadequate under MIL-F-8785C [18], which requires a value greater than 0.04. Therefore, the UAV's flight control system design must place greater emphasis on this mode.

In the lateral/directional direction, a complex-conjugate pair and two real characteristic roots are obtained, as listed in **Table 4**. Similarly, this result aligns with typical lateral/directional dynamics: one oscillatory mode, the Dutch roll, and two non-oscillatory modes, the roll subsidence and the spiral. In this case, the flight control system design needs to account for the spiral mode, which is unstable with a doubling time shorter than the recommended doubling time in MIL-F-8785C. By that standard, the spiral mode can be unstable, but only if the doubling time is more than 20 seconds for the cruising phase.

2.3. Dynamics Characteristics Derivation by Flight Testing

This section elaborates on the dynamic characteristics derived using results from a series of bMFE-UAV Flight Tests. These flight tests were conducted by the Research Center for Aeronautics Technology of the Indonesian National Research and Innovation Agency (BRIN), as shown in Figs. 8, 9, and 10.



Figure 8. Pre-Flight Preparation.



Figure 9. The bMFE-UAV Take-Off Process.

https://doi.org/10.31436/iiumej.v27i2.4013



Figure 10. The bMFE-UAV Flew Successfully.

2.3.1. Flight Test Scenario

The flight tests were conducted under the conditions specified in **Table 5**, with a maximum wind speed of 0.5 m/s. All tests were conducted manually at an altitude of around 300 meters above ground level.

Table 5. Precondition before Taking Off

Parameter at Take-Off	Value
Primary Battery Voltage	25.16 volt
Weight	7.5 kg
CG	0.25 MAC
Starting Time of Flight Test	08.31 WIB
Longitude-Latitude	7°38'59"S 107°41'32"E
Ground Elevation	9 m
Measured Temperature	21 ºC
Wind Speed	0.5 m/s

Table 6. Flight Test Scenario

No.	Mode	Input	Type	Magnitude
1	Longitudinal	Elevator	singlet	+ 15º
2	Lateral/ Directional	Aileron	doublet	+10º, then -10º

Table 7. Flight Test Procedure in Longitudinal Motion

No.	Task	Time (sec)	
1	Trim the aircraft manually and stay in that condition	10	☐
2	Provide elevator singlet input	2	☐
3	Let the aircraft fly freely longitudinally for 10 seconds	10	☐
4	Return to the "race track" pattern so the plane does not leave a controllable visual range	30	☐
	Total execution time	52	

Two test scenarios were conducted, the longitudinal test and the lateral/directional test, as listed in **Table 6**. In the former, a singlet elevator deflection is used to trigger longitudinal dynamics of the bMFE. On the other hand, to induce lateral/directional motion, a doublet aileron deflection is given. The magnitudes, however, cannot always be consistent due to the

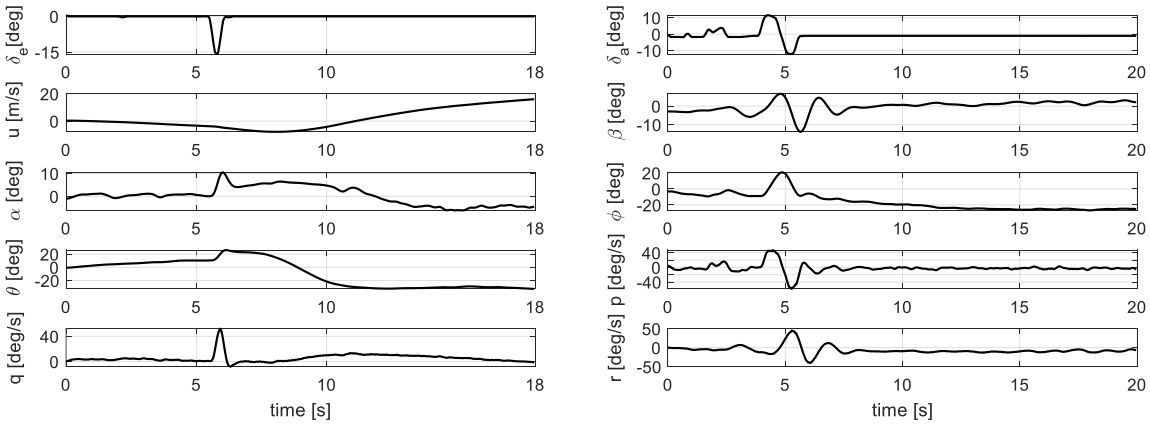
nature of pilot control. All flight data is recorded until 12 minutes after the inputs were given. The procedures for the two test scenarios are detailed in Tables 7 and 8.

Table 8. Flight Test Procedure in Lateral/Directional Motion

No.	Task	Time (sec)	
1	Trim the aircraft manually and stay in that condition	10	☐
2	Provide aileron doublet input	3	☐
3	Let the aircraft fly freely for 20 seconds	20	☐
4	Return to the "race track" pattern so the plane does not leave a controllable visual range	30	☐
	Total execution time	63	

2.3.2. Flight Data Result

Figure 11(a) presents the recorded flight data from the longitudinal mode test as deviations from the initial steady-state (trimmed) flight with an airspeed of 15 m/s. Here, the forward speed momentarily decreases after the input is given, then increases, indicating acceleration. Additionally, the angle of attack and pitch angle rise at the beginning but soon stabilize at a consistent value, while the pitch rate tends to return to zero. These results suggest that the aircraft is stable but will drift away from its initial trim condition, even with the elevator input removed.



(a) (b)

Figure 11. bMFE-UAV recorded dynamics during flight test, including the (a) longitudinal response due to elevator singlet input (impulse), and (b) lateral/directional response due to aileron doublet input.

Figure 11(b), on the other hand, shows the result of the lateral/directional flight tests of bMFE. In this case, the steady-state velocity recorded is 18 m/s. The doublet aileron inputs were not as clean as preferred, where oscillation is recorded before the intended input. This is likely due to the presence of crosswind from the left, which induced the negative sideslip angle (β) on the steady state condition. In response to the input, the UAV rolled to the right, creating a countering positive sideslip for a short period before returning to the initial trim condition. The roll rate then returned to zero, though not precisely, causing the vehicle to take slightly longer to settle its roll angle.

2.3.3. Parameter Identification

Based on the collected flight test data, a parameter identification process can be performed to obtain the aircraft's dynamic characteristics, such as natural frequency, damping ratio, and half-life. The chosen method for parameter identification is the Ordinary Least Squares (OLS)

[11, 12, 19]. Since the flight data are recorded at discrete time steps, the aircraft state-space equation for motion dynamics, as shown in Eqs. (15) and (18) are reformulated in discrete time as shown in Eq. (21).

$$\bar{z} = H\bar{\theta} + \bar{v} \tag{21}$$

where $\bar{\theta}$ is the vector of the parameters to be identified. The parameter vector estimation is calculated using the OLS error criteria in Eq. (22):

$$\bar{\theta} = [H'H]^{-1}H'\bar{z} \tag{22}$$

The estimated parameters from the OLS method are presented in **Table 9** for the longitudinal mode and **Table 10** for the lateral/directional responses. These parameters include the natural frequency, damping ratio, and half-life time. The confidence intervals and standard errors are also provided to indicate the reliability of the estimates. The composition of complex-conjugate pairs matches the typical modes of aircraft, and thus they are assigned accordingly, as shown in Tables 9 and 10.

Table 9. Dynamic characteristics of bMFE-UAV, obtained from the flight test on the longitudinal responses

Mode	Parameter	Symbol	Value
Short Period	Characteristics Roots	s_{sp}	-1.329 ± 0.582i
	Natural Frequency [rad/s]	ω_{n-sp}	1.4560
	Damping Ratio	ς_{sp}	0.9160
Phugoid	Characteristics Roots	s_{ph}	0.028 ± 0.114i
	Natural Frequency [rad/s]	ω_{n-ph}	0.1168
	Damping Ratio	ς_{ph}	-0.2368

Table 10. Dynamic characteristics of bMFE-UAV, obtained from the flight test on the lateral/directional responses

Mode	Parameter	Symbol	Value
Dutch Roll	Characteristics Roots	s_{dr}	-0.435 ± 4.221i
	Natural Frequency [rad/s]	ω_{n-dr}	4.2435
	Damping Ratio	ς_{dr}	0.1025
Roll Subsident	Characteristics Roots	s_{rs}	-1.6309
	Half-life time [s]	$T_{\frac{1}{2}spi}$	0.4250
Spiral	Characteristics Roots	s_{spi}	-0.5436
	Half-life time [s]	$T_{\frac{1}{2}spi}$	1.2752

Figure 12(a) to (b) shows the comparison between the real data and the estimated data based on the identified parameters. The comparison demonstrates that the estimated model accurately captures the aircraft's dynamic characteristics, with minimal residuals indicating a good fit between the model and the real data.

The coefficient of determination, R^2, is used to assess the proportion of variance in the dependent variable that is explained by the independent variables. A higher R^2 value indicates a better prediction model, signifying that the independent variables jointly explain a large portion of the variation in the dependent variable.

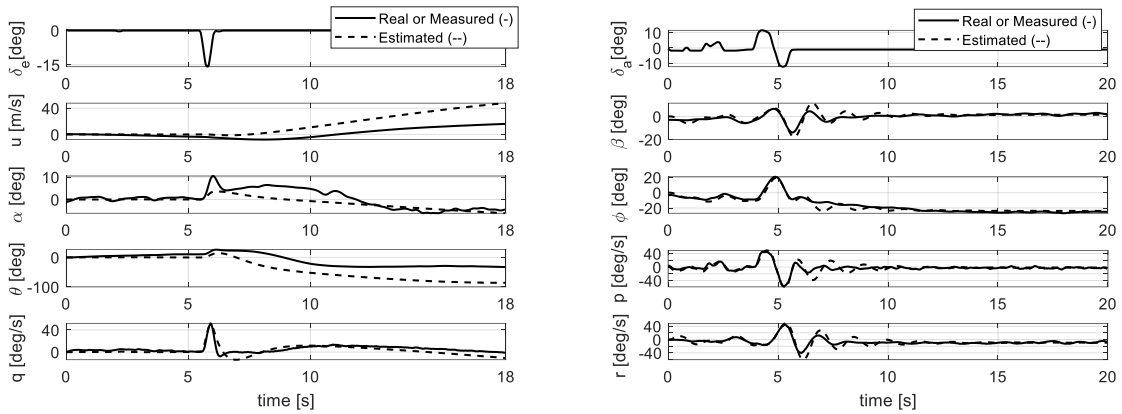


Figure 12. The comparison of real flight test results and the estimated results on (a) the longitudinal mode, and (b) the lateral/directional mode.

Conversely, a lower R^2 value suggests that the independent variables have limited explanatory power [20]. The R^2 value is calculated using the following Eqs. (23) and (24).

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} = 1 - \frac{\sum_i e_i{}^2}{\sum_i (y_i - \bar{y})^2} \tag{23}$$

where,

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \tag{24}$$

Table 11. Coefficient of determination and MSE for longitudinal motion variables

Parameter	R^2	MSE
u	0.86	0.013
α	0.72	0.016
θ	0.85	0.017
q	0.66	0.004

Table 12. Coefficient of determination and MSE for lateral/directional motion variables

Parameter	R^2	MSE
β	0.74	0.006
φ	0.93	0.003
p	0.72	0.004
r	0.80	0.003

After calculating the coefficient of determination and MSE, all included variables are considered acceptable, with the pitching rate variables for longitudinal dynamics and the rolling rate for lateral/directional dynamics yielding the lowest values. These indicate that the proposed model, derived from the flight test data, effectively captures a significant portion of the variance in the dependent variables, using the selected independent variables. The lowest MSE indicates that the average prediction error is small.

3. Results Comparison and Analysis

Tables 13 and 14 compare the results of bMFE's dynamic characteristics derived from analytical calculations and flight-test data identification.

3.1. Overall Comparison

Generally, the analytical method is expected to offer a clear, systematic approach grounded in theory and to provide a strong foundation for the final design and stability analysis. However, it may not account for all real-world variables and disturbances. On the other hand, the flight test combined with the OLS identification method leverages actual flight data, capturing real-world dynamics but potentially introducing noise and experimental errors.

Table 13. The difference in longitudinal dynamic characteristics between the analytical and the flight test method

Parameters		Results		Difference (from analytical)
		Analytical	Flight Data	
Short Period	s_{sp}	-3.711 ± 0.710i	-1.329 ± 0.582i	-
	ω_{n-sp}	3.7785 /s	1.4560 /s	61.4%
	ς_{sp}	0.9822	0.9160	6.1%
Phugoid	s_{ph}	-0.026 ± 0.736i	0.028 ± 0.114i	-
	ω_{n-ph}	0.7368 /s	0.1168 /s	84.1%
	ς_{ph}	0.0356	-0.2368	-

Table 14. The difference in lateral/directional dynamic characteristics between the analytical and the flight test method

Parameters		Results		Difference (from analytical)
		Analytical	Flight Data	
Dutch Roll	s_{dr}	-5.053 ± 10.30i	-0.435 ± 4.221i	-
	ω_{n-dr}	11.47 /s	4.244 /s	63.0%
	ς_{dr}	0.445	0.103	76.4%
Roll Subsidence	s_{rs}	-16.557	-1.6309	-
	$T_{\frac{1}{2}_{rs}}$	0.0419 s	0.4250 s	914.3%
Spiral	s_{spi}	0.053	-0.544	-
	$T_{\frac{1}{2}_{spi}}$	(13.1825 s)*	1.2752 s	-

* Doubling-time

As listed in Tables 13 and 14, both methods consistently yielded oscillatory behavior, enabling the assignment of typical aircraft modes. In the longitudinal mode, two oscillatory modes are observed, as indicated by the two complex-conjugate pairs. In the lateral/directional case, both methods yielded complex-conjugate pairs together with two real roots. This manifested in the lateral/directional time response as a single oscillatory motion, accompanied by two non-oscillatory components (see **Figure 12**).

Evidently, only the damping-ratio results for the Short-period mode show close agreement, with a difference of less than 10%. Others, including the Short Period's natural frequency, displayed significant discrepancies between the methods' results. However, the flight test method consistently gives a lower oscillation frequency and a longer half-life. The lack of an actuator dynamic model in the analytical method calculation might cause these. Actuator dynamics should introduce delays and phase lag, which can reduce the natural frequency, as shown by a shift in the real parts of the characteristic roots to the left of the complex plane. The damping ratio, especially for the Short-period mode, might not change as much because it is more affected by aerodynamic factors. Other reasons for these discrepancies can include

https://doi.org/10.31436/iiumej.v27i2.4013

inaccuracies in mass and inertia distribution, the lack of a structural flexibility model, or nonlinear effects.

Moreover, the discrepancies shown are not limited to the value of dynamic parameters but also to stability. For instance, the results in the Phugoid mode indicate that the analytical method predicts the UAV is stable, whereas the flight test indicates instability. **Figure 12**(a) shows that instability happens when the vehicle's forward speed, which is a dominant factor for phugoid mode, increases by more than 40 m/s from the initial 15 m/s airspeed. In this condition, the assumption of linearity of the aerodynamic derivative cannot hold, leading to stability inversions.

On the other hand, the spiral mode produced a different result: the analytical method yielded an unstable result, whereas the flight data showed stability. This result indicates that some unknown damping factors may not yet have been accounted for in the analytical model. Another reason is that since the damping in the rolling and yawing rates is higher in the flight test result, the roll angle range is much more limited, only from -20º to 20º. In contrast, the roll result is nearly 60 ° to the right in the analytical method. This difference in roll envelope may contribute to the discrepancies observed in the lateral/directional stability results.

3.2. Steps in Derivation Process

To derive the dynamic characteristics of the bMFE UAV, the analytical approach requires significantly more steps than the flight test method. In modeling, each factor and its derivatives with respect to the six motion parameters need to be derived. Aerodynamics alone requires at least twelve static and dynamic derivatives, not to mention the control derivatives. Using software like the USAF DATCOM can reduce the time required, but only if all aircraft configurations and flight conditions are well-defined. The process of finding the Center of Gravity and Inertia can also pose a challenge for an off-the-shelf UAV, as it often requires numerous experiments.

Additionally, deriving the propulsion effects for an already installed system is complex. The more accurate the model needs to be, the more effort is required to determine each factor. However, from the facility's point of view, all this can be done in a laboratory using a single good computer. On the other hand, flight testing coupled with an identification process is more straightforward. Flight testing for an off-the-shelf small UAV can be conducted immediately, yielding enough data to support an identification process. Moreover, the results can represent the entire configuration and conditions during the flight test in a single all-in-one experiment. The most essential requirement is an experienced pilot who can establish an initial trim condition, apply the input correctly, and let the plane fly freely to be recorded, while avoiding collisions and crashes. However, this method also requires a flight test field and fair and consistent weather conditions.

Nevertheless, for a small UAV, this research has shown that it is more practical and efficient to focus on the flight-test and identification approach for deriving dynamic characteristics. Small UAVs such as the bMFE-UAV can be quickly deployed and flown, making it easier to gather flight-test data without extensive preparatory work. The flexibility to repeat flight tests under different conditions also facilitates the design of robust control systems capable of handling various scenarios.

3.3. Implications in Designing the Automatic Flight Control System for Small UAVs

The comparative analysis between the analytical and flight test methods highlights the practical advantages of the latter approach. For small UAVs already built and available for

immediate use, the flight test method offers a straightforward way to identify the vehicle's dynamic characteristics. As the UAV is flown repeatedly and the flight test data is consistently obtained, the accuracy of the derived characteristics improves over time. This tendency implies that control system design, such as pitch hold, altitude hold, and stability augmentation, can be further refined through additional flight tests.

Some concerns exist regarding differences in dynamic parameters obtained from the flight-test method compared with the analytical approach. These discrepancies may warrant further investigation. However, despite these differences, the flight test method remains more practical for designing flight control systems. Ensuring that analytical and flight test results align precisely may not be as crucial as the immediate benefits of the flight test method for practical control system design.

The flight test method is a part of the data-driven approach, which is not only becoming increasingly common but also more viable and promising due to advancements in sensor technology and data processing capabilities. This approach leverages real-time data to provide accurate and up-to-date insights into the UAV's performance. As sensors become more sophisticated and accessible, the ability to gather detailed flight data has improved, making the data-driven approach more viable and effective for UAV control system design. This method is particularly suitable for small fixed-wing UAVs, which have also become more accessible and cheaper, further enhancing the practicality and appeal of the data-driven approach.

4. CONCLUSIONS

UAV dynamic parameters for a small fixed-wing UAV, the bMFE-UAV, have been derived using two approaches. The first approach is analytical, utilizing a mathematical model derived from physical principles with data from the USAF DATCOM. The second approach involves flight testing and data analysis using Ordinary Least Squares (OLS). The results include parameters such as damping ratio and natural frequency, from which flight modes can be identified. These results have been compared, and both methods have yielded consistent oscillatory behavior, enabling the assignment of typical aircraft modes. Only the short-period mode shows close agreement, with a difference of less than 10%. Others, including the short-period's natural frequency, displayed significant discrepancies between the method's results. The observed discrepancies underscore the inherent limitations of analytical methods in capturing the complexity of real flight dynamics. Accordingly, this research provides empirical evidence that flight testing remains indispensable, as it produces reliable data to support the design, validation, and fine-tuning of small UAV flight control systems.

ACKNOWLEDGMENT

The research was supported by Research Center for Aeronautics Technology, National Research and Innovation Agency, Indonesia and was conducted as part of the Development of Flight Control and Guidance Systems for Advanced Vehicles research, under the 2021 - Penelitian, Pengabdian Masyarakat dan Inovasi (P2MI-2021) program of Flight Physics Research Group, Faculty of Mechanical and Aerospace Engineering, Institut Teknologi Bandung.

REFERENCES

[1] Dimitriadis G., Panagiotou P., Dimopoulos T., & Yakinthos K. (2024). Aerodynamic Stability Derivative Calculations Using the Compressible Source and Doublet Panel Method. Journal of Aircraft, 61(4), 1034-1046. https://doi.org/10.2514/1.C037747.

[2] Okhovatian S., & Koukounian V. (2024). Using Empirical Data to Validate the Role of Computational Fluid Dynamics in Various Stages of Aero-Acoustic Simulations. Canadian Acoustics - Acoustique Canadienne, 52(1). https://doi.org/10.1121/10.0026809.

[3] Liu Y. (2023). Harnessing CFD Modeling Techniques for Optimizing Automotive Streamlines and Reducing Aerodynamic Drag. Highlights in Science, Engineering and Technology. https://doi.org/10.54097/v25ka918.

[4] Olejnik A., Dziubiński A., & Kiszkowiak Ł. (2021). Reliable method of aerodynamic analysis using computational fluid dynamics and scaled models in the development process of a Very Light Airplane. IOP Conference Series: Materials Science and Engineering (1), 1024. https://doi.org/10.1088/1757-899X/1024/1/012048.

[5] Tai S., Bu C., Wang Y., Yue T., Liu H., & Wang L. (2024). Identification of aircraft longitudinal aerodynamic parameters using an online corrective test for wind tunnel virtual flight. Chinese Journal of Aeronautics, 37(9), 261-275. https://doi.org/10.1016/j.cja.2024.05.031.

[6] Corneliu S., Burghiu M., Trandafir E., Apostol E., Nica A., Pana A., Balasa R., Curt D., Defta S., Vajaiac E., Pirvu C. (2023). Wind Tunnel Testing of a Common Research Model. INCAS Research Challenges in Aerospace Technologies. https://doi.org/10.2514/6.2023-1017

[7] Benyamen H., Benjamin S. Mays. Chowdhury M., Keshmiri S., & Mark S. (2023). Analysis of Aircraft Simulation Validity in Different Flight Conditions. International Conference on Unmanned Aircraft Systems. https://doi.org/10.1109/ICUAS57906.2023.10156586.

[8] Xu S., Bi W., Zhang A., & Mao Z. (2022). Optimization of flight test tasks allocation and sequencing using genetic algorithm. Applied Soft Computing, 115. https://doi.org/10.1016/j.asoc.2021.108241.

[9] Weishaeupl A. B., Scanlan J., & Sobester A. (2023). Flight Test Driven Development of Low Cost UAVs - Pitfalls and Opportunities. AIAA 2023-1739. https://doi.org/10.2514/6.2023-1739.

[10] Mohd Shah H. N., Mohamad Sebir M. A., Abdollah M. F., Baharon M. R., Ahmad A., & Ali Arshad M. (2023). Develop and Design Small Scale UAV. Modern Applied Science, 17(2), 49. https://doi.org/10.5539/mas.v17n2p49.

[11] Jategaonkar R. V. (2006). Flight Vehicle System Identification (978-1-62410-279-0 ed.).

[12] Klein V., & Morelli E. A. (2006). Aircraft System Identification: Theory and Practice.

[13] Wang L., Zhao R., & Zhang Y. (2022). Aircraft Lateral-Directional Aerodynamic Parameter Identification and Solution Method Using Segmented Adaptation of Identification Model and Flight Test Data. Aerospace, 9(8). https://doi.org/10.3390/aerospace9080433.

[14] Ali S., Hassan O., Gopalakrishnan A., Muriyan A., & Francis S. (2022). Unmanned Aerial Vehicles: A Literature Review. Hunan Daxue Xuebao/Journal of Hunan University Natural Sciences, 49(7), 96-113. https://doi.org/10.55463/issn.1674-2974.49.7.11.

[15] Dias J. N. (2023). Flight-Test Determination of Longitudinal Stability Using System Identification. Journal of Aircraft, 60(5), 1659-1674. https://doi.org/10.2514/1.C037252.

[16] Fighter Hand launch fixed wing [https://diydrones.com/profiles/blogs/fighter-hand-throw-fixed-wing].

[17] Roskam J. (2001). Airplane Flight Dynamics and Automatic Flight Controls - Part 1.

[18] Departments and Agencies of the Department of Defense. (1980). Military Specification Flying Qualities of Piloted Airlines.

[19] Li H. (2024). Aerodynamic parameter identification method of unmanned aerial vehicles (UAVs), (p. 115). https://doi.org/10.1117/12.3039610.

[20] 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

https://doi.org/10.31436/iiumej.v27i2.4062

Impact of Aggressive and Moderate Driving Intensity on Vehicle Air–Fuel Ratio Stability

MOHAMAD NORSYAFIQ AB HAMID[1], MUHAMMAD YUSRI ISMAIL[1,2*], NAJMI HAZIQ BADRULHISAM[1,2,3,4]

[1]*Faculty of Mechanical and Automotive Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, Pekan, Malaysia*
[2]*Automotive Engineering Centre, Universiti Malaysia Pahang Al-Sultan Abdullah, Pekan, Malaysia*
[3]*Centre for Sustainable Advanced Engineering Design, Faculty of Engineering, Built Environment, and Information Technology, SEGi University, Selangor, Malaysia*
[4]*Faculty of Engineering, Built Environment, and Information Technology, SEGi University, Selangor, Malaysia*

**Corresponding author: yusriismail@umpsa.edu.my*

(Received: 24 November 2025; Accepted: 24 February 2026; Published online: 10 May 2026)

ABSTRACT: Aggressive driving behavior occurs when drivers neglect traffic rules, leading to violations, accidents, and increased road hazards. Such behavior endangers the driver and places other road users at risk. In contrast, moderate driving reflects disciplined patterns that reduce risks and improve vehicle efficiency. This study examines the influence of aggressive and moderate driving behaviors on engine performance, with a focus on air–fuel ratio (AFR) stability. The investigation used On-Board Diagnostics (OBD-II) data collected during 11.7 km driving sessions under two conditions: aggressive and moderate. The test vehicle, equipped with a 1.8-liter engine, was driven at controlled speeds. Aggressive driving was defined as driving in the 60–90 km/h range, while moderate driving occurred below 60 km/h, representing typical urban operation. Results show distinct differences in AFR behavior. Aggressive driving produced greater instability, with richer AFR outliers near 12.5, higher global wavelet spectrum (GWS) peaks of about 40 units compared to 20, and longer time in fuel-rich regions (12–17% versus 5–14%). These fluctuations indicate unstable combustion, increased fuel use, and higher emissions. In contrast, moderate driving maintained a more stable AFR distribution, with a greater proportion of leaner readings (~36% versus 30%) and closer alignment with stoichiometric values. This stability supports improved combustion efficiency, reduced emissions, and better fuel economy. Overall, the findings show that driving intensity significantly influences AFR stability, with moderate driving promoting cleaner and more efficient engine operation.

ABSTRAK: Tingkah laku pemanduan agresif berlaku apabila pemandu mengabai peraturan lalu lintas, menyebabkan pelanggaran, kemalangan dan peningkatan bahaya di jalan raya. Tingkah laku sedemikian membahayakan pemandu dan membahayakan pengguna jalan raya lain. Sebaliknya, pemanduan sederhana mengurangkan risiko dan meningkatkan kecekapan kenderaan. Kajian ini mengkaji pengaruh tingkah laku pemanduan agresif dan sederhana terhadap prestasi enjin, dengan memberi tumpuan kepada kestabilan nisbah udara-bahan api (AFR). Kajian ini menggunakan data Diagnostik Atas Papan (OBD-II) yang dikumpul semasa sesi pemanduan 11.7 km di bawah dua keadaan: agresif dan sederhana. Kenderaan ujian, dilengkapi enjin 1.8 liter, dipandu pada kelajuan terkawal. Pemanduan agresif ditakrifkan pada julat 60–90 km/j, manakala pemanduan sederhana berlaku di bawah 60 km/j, mewakili pemanduan tipikal dalam bandar. Dapatan kajian menunjukkan perbezaan ketara dalam tingkah laku AFR. Pemanduan agresif menghasilkan ketidakstabilan terbesar, dengan AFR

terpisah yang lebih tinggi iaitu hampir 12.5, puncak spektrum gelombang global (GWS), iaitu 40 unit lebih tinggi berbanding 20, dan masa lebih lama di kawasan kaya bahan api (12–17% berbanding 5–14%). Turun naik ini menunjukkan pembakaran yang tidak stabil, peningkatan penggunaan bahan api dan pelepasan karbon lebih tinggi. Sebaliknya, pemanduan sederhana mengekalkan taburan AFR lebih stabil, dengan kadar bacaan lebih rendah (~36% berbanding 30%) dan penjajaran hampir pada nilai stoikiometrik. Kestabilan ini menyokong kecekapan pembakaran yang lebih baik, mengurangkan pelepasan dan penjimatan bahan api yang lebih baik. Secara keseluruhan, dapatan kajian menunjukkan bahawa pemanduan yang baik mempengaruhi kestabilan AFR dengan ketara. Pemanduan sederhana menggalakkan operasi enjin yang lebih bersih dan cekap.

KEYWORDS: *Aggressive driving behavior; Moderate driving behavior; Air–fuel ratio (AFR) stability; On-Board Diagnostics (OBD II); Driving styles.*

1. INTRODUCTION

All modern vehicles, whether powered by an internal combustion engine, hybrid systems, or hydrogen, rely on braking systems [1]. The braking system is essential for slowing the vehicle down and preventing collisions [2]. In vehicles equipped with internal combustion engines, braking events are not only mechanical actions but also induce transient engine conditions, particularly during throttle release and reapplication. During deceleration, changes in throttle position and engine load trigger fuel cut-off or enrichment strategies governed by the engine control unit (ECU), which directly affect the air–fuel ratio (AFR). This is due to the working principle of the internal combustion engine, which depends on the throttle position and fuel injection system strategies [3]. These situations will contribute to transient engine conditions that can destabilize the air-fuel ratio, disrupting combustion efficiency and emissions control. These rapid transitions between fuel cut-off, idle control, and acceleration introduce AFR instability, potentially degrading combustion efficiency and emissions performance. Such transient engine behavior is strongly influenced by driver input, especially during abrupt braking and subsequent acceleration [5]. Consequently, driving style plays a critical role in determining AFR response characteristics and overall engine feedback during real-world driving conditions [4].

Driving behavior can be divided into two types: aggressive and moderate. In the event of aggressive driving, the driver tends to maximize the use of fuel due to the rapid acceleration [5]. This also resulted in frequent hard braking, which leads to danger in driving road safety, as well as higher exhaust emissions, which also resulted in inconsistent throttle input [6]. On the other hand, a moderate driving style is associated with smooth throttle control and steady speeds, moving vehicles, which leads to less gradual braking, thus a safer driving style [7]. During aggressive driving, the air-fuel ratio often exhibits rich spikes and inefficient combustion, whereas moderate driving has a lean air-fuel ratio, resulting in a better fuel-air mixture to maintain stoichiometric conditions. Meanwhile, moderate driving is a better cost-effective driving style that contributes to lower exhaust emissions as well [8]. Under aggressive handling conditions, the engine control unit (ECU) often struggles to maintain a stable air-fuel ratio, leading to ineffective combustion. In contrast, moderate driving consistently maintains a more stable air-fuel ratio, improving fuel efficiency and lowering emissions.

Aggressive and moderate driving are two common types of vehicle handling observed on the road, often depending on the driver's situation and environment. Understanding the effects of these driving behaviors on vehicle performance is essential, particularly for improving road safety under real-time or transient driving conditions [9]. From a combustion standpoint,

transient driving events are important because they directly affect the air–fuel ratio (AFR), which governs combustion stability, efficiency, and emission formation. However, most existing studies do not clearly link aggressive and moderate driving behaviors to variations in AFR during real-world on-road driving. As a result, there is still a limited understanding of the influence of different driving styles on AFR behavior beyond steady-state conditions.

Kumar and Shen [10] investigated a discrete-time, model-based estimation and control approach for cycle-to-cycle air–fuel ratio (AFR) regulation in spark-ignition engines under transient operation, explicitly addressing combustion imbalance caused by rapid throttle changes. Experimental validation on a full-scale gasoline engine test bench demonstrated that the estimated AFR closely matched lambda-sensor measurements, with errors generally below 3.1% across engine speeds of 1000, 1500, and 2000 $r \cdot min^{-1}$ and engine loads of 60, 120, and 180 N·m, confirming improved mixture preparation, enhanced combustion stability, and reduced cycle-to-cycle combustion variability under varying operating conditions.

Daham et al. [11] tested driver variability using a fully warmed-up Euro 2 spark-ignition Ford Mondeo operated on a repeated 0.6 km urban driving loop with multiple junction events. Although mean vehicle speed varied only slightly, large differences in acceleration, throttle position, and throttle jerk were observed among drivers, indicating fundamentally different transient driving behaviors. These rapid throttle transients were shown to disturb the $\lambda = 1$ air–fuel ratio control system, leading to AFR instability and increased cycle-to-cycle combustion variability during real-world driving, particularly under aggressive throttle inputs.

Rimpas D et al. [12] proposed a vehicle monitoring framework that collects key operational parameters such as vehicle speed, engine load, airflow, and fuel consumption through electronic sensors and retrieves them through the OBD-II diagnostic protocol. Experimental validation was conducted over a 5 km urban route under both light and heavy traffic conditions, demonstrating stable and reliable data acquisition throughout the test. Quantitative analysis revealed that aggressive driving behavior, characterized by frequent acceleration and deceleration events, substantially increased fuel consumption, whereas steady driving conditions yielded more consistent engine load profiles and improved fuel efficiency. Notably, the results indicated that vehicles exhibited lower fuel consumption during steady cruising at higher speeds, contradicting the common assumption of uniformly higher fuel consumption at higher speeds. These findings confirm the capability of OBD-based monitoring systems to quantitatively assess the influence of driving behavior on vehicle performance and fuel efficiency.

Zeb et al. [13] developed an OBD-II-based cyber-physical system (CPS) for identifying road bottlenecks and quantifying their associated inefficiencies in terms of fuel consumption, travel time, and CO_2 emissions. The proposed system integrates vehicular parameters, including vehicle speed, engine RPM, mass airflow (MAF), intake air temperature (IAT), and air–fuel ratio (AFR) , acquired via OBD-II, alongside GPS and temporal data transmitted via Wi-Fi to a cloud analytics platform. The solution was experimentally validated on a fixed urban route over five consecutive days, enabling consistent comparison across traffic conditions. Quantitative analysis revealed that two bottlenecks during morning trips accounted for 38 % of total fuel consumption, 31 % of CO_2 emissions, and 31 % of overall travel time, while three evening bottlenecks contributed to 51 % of fuel consumption, 50 % of CO_2 emissions, and 29 % of trip duration. Instantaneous fuel consumption profiles further demonstrated pronounced spikes during low-speed, stop-and-go conditions, confirming the strong coupling between traffic congestion and fuel inefficiency. The study demonstrated that OBD-II-driven CPS frameworks can effectively localize road bottlenecks, quantify their operational costs, and

provide valuable data for traffic model calibration and intelligent transportation system development.

Padmanaban et al. [14] compared aggressive and non-aggressive driving behaviors using an on-road dataset and EPA fuel economy cycles, focusing on acceleration patterns, fuel consumption, and environmental impact. Results showed that aggressive driving involved short, intense accelerations linked to higher fuel use and potential vehicle wear, while non-aggressive driving featured smoother, sustained accelerations, supporting better fuel economy and reduced environmental impact.

Li et al. [15] proposed a representing-behavior-based driving style recognition (RB-DSR) method capable of identifying and tracking dynamic changes in driving behavior. The model integrates a driving-style distinction metric with k-means clustering and a genetic algorithm-based feature selection (GA-FS) to optimize representative-behavior recognition. Road tests confirmed its effectiveness for real-time driving style analysis, though the authors noted the need for larger datasets to improve robustness. Muwaffaq et al. [16] examined the relationship between risky driving behaviors and anger expression among Nigerian drivers using Multilevel Latent Class Analysis (MLCA). The study found that younger male drivers were more prone to aggressive anger expression and recommended improved driver training and stricter traffic law enforcement to enhance road safety.

ŚLUSARCZYK and Jurecki [17] developed a WEco index using a driving simulator to assess eco-driving behavior. Testing 37 drivers, the study identified three performance levels, showing significant differences in fuel use and engine RPM, supporting its use in driver training and fleet management. Zhang et al. [18] proposed an eco-driving evaluation method using speed-specific Vehicle-Specific Power (VSP) distributions from 19,779 drivers. The approach established VSP and fuel-rate baselines, revealing a fuel-consumption difference of over 20% between eco and aggressive driving behaviors.

Adavikottu and Velaga [19] analyzed the impact of Aggressive Driving Behavior (ADB) on collision risk using a driving simulator with 58 drivers across aggression levels. Results showed that aggressive drivers reduced time-to-collision (TTC) and speed-reduction time (SRT) by 82% and 38%, respectively, leading to a much higher crash risk. The study highlights that delayed braking and higher approach speeds among aggressive drivers significantly increase the likelihood of accidents. Adavikottu et al. [20] found that aggressive drivers showed 25% lower speed variability, 18% shorter recovery time, and 107% more tailgating, with a 23.77% higher crash risk during car-following. The study underscores the need for safer car-following models in automated driving systems. Ma and Zhang [21] found that aggressive and moderate drivers exhibited greater anxiety and aggression around automated vehicles than around human-driven ones, while defensive drivers showed no change. The study offers insights into improving safety in mixed-traffic systems. Adavikottu and Velaga [22] analyzed the impact of driving aggression on lane change behavior. Aggressive drivers changed lanes up to 59% faster and showed over five times higher crash risk than non-aggressive drivers, underscoring the need for adaptive driver-assistance systems to enhance safety. Kerwin and Bushman [23] examined public perceptions of aggressive driving using videos depicting close following, illegal passing, and near-collision scenarios viewed from different perspectives. Findings showed that tailgating and illegal passing were rated highly aggressive, particularly from a third-person perspective, whereas high speeds alone did not strongly influence perceptions. The study highlights that perceived aggressiveness depends on viewing perspective and individual traits such as anger and driving attitudes.

https://doi.org/10.31436/iiumej.v27i2.4062

This study aims to investigate the relationship between driving behavior and air–fuel ratio (AFR) stability under two distinct driving styles, namely aggressive and moderate driving, using real-time On-Board Diagnostics (OBD-II) data. The experimental tests were conducted along an 11.7 km driving route with speed variations ranging from below 60 km/h (moderate) to 60–90 km/h (aggressive), representing typical urban driving conditions. The focus on AFR behavior under these conditions was emphasized for several reasons: first, AFR stability plays a critical role in ensuring efficient combustion and fuel economy, particularly in transient driving situations; second, real-time AFR response under dynamic throttle conditions remains underexplored, despite its direct relevance to driver behavior and vehicle control systems. To date, few studies have examined AFR variations in response to driver-induced acceleration and braking behavior, with most prior research emphasizing emission or fuel-consumption patterns rather than combustion balance. Moreover, studies of AFR dynamics often focus on laboratory or chassis-dynamometer settings, lacking real-road driving data that reflect natural driving tendencies. Therefore, this study addresses an existing gap by analyzing AFR fluctuations and throttle response behavior under real-world driving conditions, offering insights into how aggressive and moderate driving styles influence combustion stability. The novelty of this study lies in integrating driving behavior analysis with AFR stability assessment via OBD-II monitoring, providing a cost-effective, data-driven method for understanding driver–engine interaction. By quantifying AFR deviations and identifying patterns of fuel-rich and lean operation zones, the study contributes to a deeper understanding of how driving style affects fuel efficiency and engine stability. Furthermore, these findings offer practical implications for developing eco-driving frameworks, adaptive vehicle control algorithms, and driver behavior recognition systems to enhance energy efficiency in real-world conditions.

2. METHODOLOGY

A passenger car compliant with OBD-II/EOBD (CAN, ISO-15765; optionally ISO-9141-2/ISO-14230) was used to record vehicle data under driving conditions. The test vehicle used in this study was a passenger car equipped with a 1.8-liter gasoline spark-ignition V6 engine and an automatic transmission. All experiments were conducted under fully warmed-up conditions during urban driving operation. The details of the vehicle are shown in Table 1.

Table 1. Vehicle specification

Parameter	Description
Vehicle category	Passenger car
Model year	2008
Engine configuration	V6, naturally aspirated
Engine displacement	1.8 L
Fuel type	Gasoline
Transmission type	Automatic transmission
Emission standard	Euro 2
Engine operating condition	Fully warmed-up

The vehicle diagnostic data was acquired using a professional OBD-II diagnostic scanner. The data was captured at a sampling rate of 1 Hz. Each driving test was conducted for approximately 15–20 minutes, resulting in a representative dataset for both aggressive and moderate driving styles. For each driving style, three repeated runs were performed to ensure data consistency and repeatability. All experiments were carried out under similar ambient conditions, with temperatures ranging from 28 to 32 °C, dry weather, and normal traffic flow. To minimize driver-related variability, the same driver was used for both aggressive and

https://doi.org/10.31436/iiumej.v27i2.4062

moderate driving tests, and all runs were conducted on the same route using the same vehicle. All driving experiments were conducted under comparable driving conditions to minimize external influences on vehicle operation. The tests were carried out on the same predefined route and were completed on the same day within a limited time window. Road conditions are mostly urban conditions. Weather conditions were stable, with no rainfall and ambient temperatures remaining within a narrow range. Traffic conditions were moderate and consistent, with no abnormal congestion or incidents observed during the test runs. By maintaining similar roads, weather, and traffic conditions, variations in the recorded data are primarily attributed to differences in driving style rather than environmental factors.

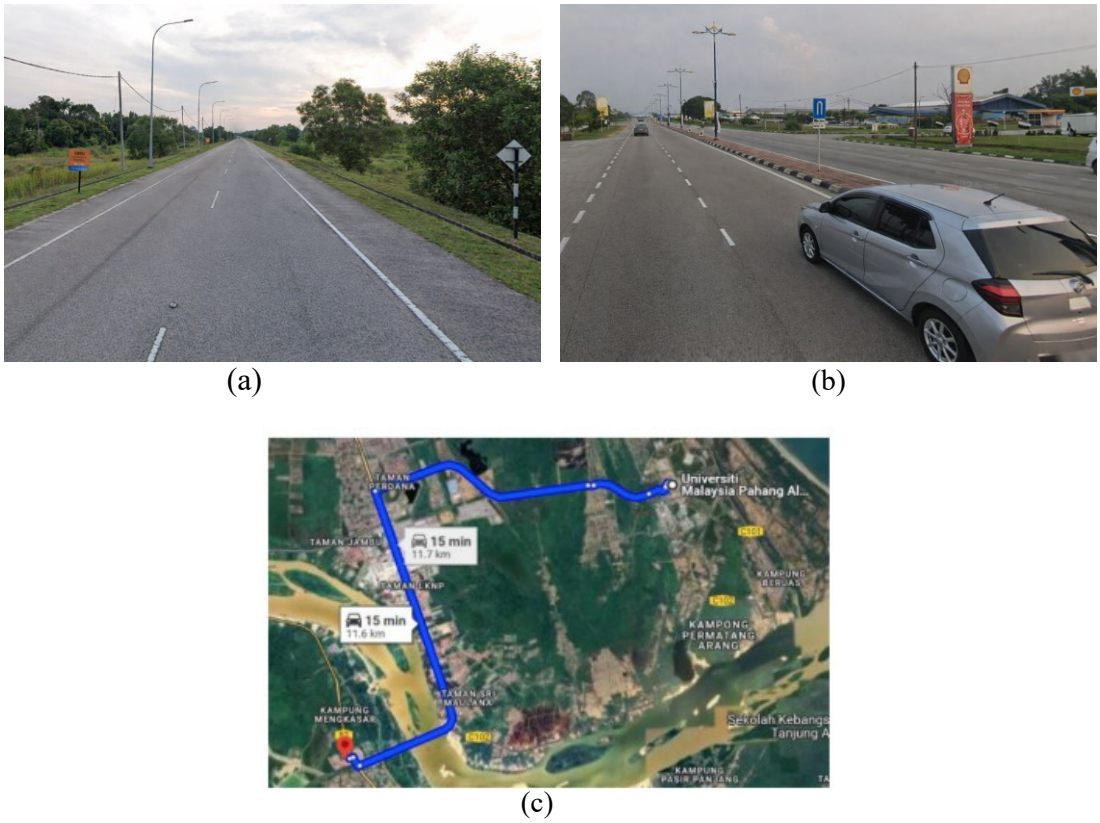


Figure 1. Selected test routes consisting of (a) a suburban single-lane road and (b) an urban dual-lane road, both driven in forward and reverse directions, and (c) the Google Maps overview of the complete test route (15–20 min) used for the diagnostic scanner experiment.

During the experiment, the scanner's Vehicle Communication Interface (VCI) was plugged into the OBD-II port to enable real-time data communication (engine speed, air–fuel ratio, coolant temperature, throttle position) via the CAN bus/ISO protocols, facilitating the retrieval of diagnostics from the ECU. The vehicle was driven along a designated 11.7 km route designed to reflect real-world driving conditions, as shown in Figure 1. The journey included a balanced mix of acceleration, steady cruising, deceleration, and idling phases. Throughout the test, the driver operated the vehicle naturally without any imposed controls or restrictions, ensuring that the dataset captured authentic variations in driving behavior. Each AFR reading represents a single snapshot taken during this representative driving session. Each AFR data point corresponds to a snapshot within this driving session. Air–fuel ratio (AFR) measurements were collected at high frequency, resulting in several hundred data points over the 11.7 km driving route. To maintain accuracy, invalid or spurious readings those falling outside the

sensor's physical limits were excluded from the dataset. For consistency, AFR values greater than 17.5 were capped at 17.5, as readings above this threshold typically correspond to deceleration fuel cut-off (DFCO) events. On the lower end, no AFR values below 12.0 were recorded during this drive. The AFR readings were organized into five fixed intervals for analysis: 12.0–13.1, 13.1–14.2, 14.2–15.3, 15.3–16.4, and 16.4–17.5. The flow chart of the experiment. The flow chart for data recording and analysis is shown in **Figure 3**. The vehicle details are in **Table 1**.

2.1. Criteria for Classification of Aggressive and Moderate Driving Behaviors

The driving conditions are summarized in Table 2. Two driving styles, namely moderate and aggressive, were evaluated under the same route configuration to ensure a fair comparison. For each driving style, the test was repeated three times, with a target duration of 15–20 minutes per run. Moderate driving was conducted at vehicle speeds ranging from 50 to 80 km/h, representing relatively steady driving with minimal braking, typically observed during morning travel after the peak rush hour. In contrast, aggressive driving was characterized by greater speed variations between 80 and 100 km/h, reflecting unsteady driving behavior with frequent acceleration and deceleration due to traffic interactions and slowing-down events. This experimental design allows the assessment of vehicle response and AFR behavior under both steady and transient real-world driving conditions. There is no clear literature that defines aggressive and moderate driving. But these two features are quite common in the investigation of vehicle analysis behavior. A study by Adavikottu and Velaga [19] reported mean vehicle speeds for non-aggressive, moderately aggressive, and aggressive drivers as 40.4 km/h, 49.8 km/h, and 53.7 km/h, respectively. Aggressive drivers maintained speeds that were 13.3 km/h higher than those of non-aggressive drivers and 3.9 km/h higher than those of moderately aggressive drivers, indicating a tendency to sustain higher speeds. Meanwhile, in another study by Kerwin and Bushman [23], tailgating scenarios were evaluated at vehicle speeds of 35 mph (56 km/h) and 65 mph (105 km/h), representing typical urban and highway driving conditions. The degree of aggressiveness was further assessed using Time-to-Collision (TTC), which represents the time required for the following vehicle to collide with the leading vehicle if it were to stop suddenly.

Table 2. Summary of driving styles and test conditions for the urban–suburban mixed driving route

Driving style	Route type (mixed driving route)	Runs (repeat)	Target duration per run	Variable driving speed (km/h)	Time
Moderate	Urban–suburban	3	15–20 min	50 – 80 (steady driving with minimum braking)	Morning after rush hour, going to work
Aggressive	Urban–suburban	3	15–20 min	80 – 100 (unsteady driving due to slowing down)	

2.2. Wavelet Power Spectrum Analysis

Wavelet transform analysis is widely used to evaluate variations in non-stationary time-series signals, particularly under transient operating conditions. In this study, wavelet analysis is applied to air–fuel ratio (AFR) time-series data to investigate temporal and frequency-dependent fluctuations arising from real-world driving behavior. Through this approach, AFR variability can be decomposed simultaneously in both the time and frequency domains. A wavelet can be expressed as a function that has a zero mean and finite energy. A continuous wavelet transform (CWT) for a wavelet function can be expressed as Equation (1) [24]

https://doi.org/10.31436/iiumej.v27i2.4062

$$CWT(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t)\psi^* \left(\frac{t-b}{a}\right), \qquad a,b \in R, a \neq 0. \tag{1}$$

To analyze transient AFR variations, the wavelet power spectrum (WPS) and global wavelet spectrum (GWS) are used. The squared magnitude of the continuous wavelet transform (CWT) represents the energy of the AFR signal at a given scale and time. The WPS is obtained by normalizing the wavelet power by the scale parameter, which allows the AFR energy to be compared with a white-noise background [25]. Accordingly, the AFR data are normalized using Equation (2).

$$WPS_n = \frac{|CWT_n(a)|^2}{\sigma^2} \tag{2}$$

GWS is described as a time average and denoted as, W_s, peak location in GWS is the highest periodicity at a particular given time series. Therefore, the formula is as Equation (3).

$$GWS = W_s = \frac{1}{N} \sum_{n=1}^{N} |CWT(a)|^2 \tag{3}$$

Through the combined use of WPS and GWS, the transient characteristics and dominant frequencies of AFR variation under different driving conditions can be effectively identified. Data programming through the MATLAB interface is shown in Figure 2.

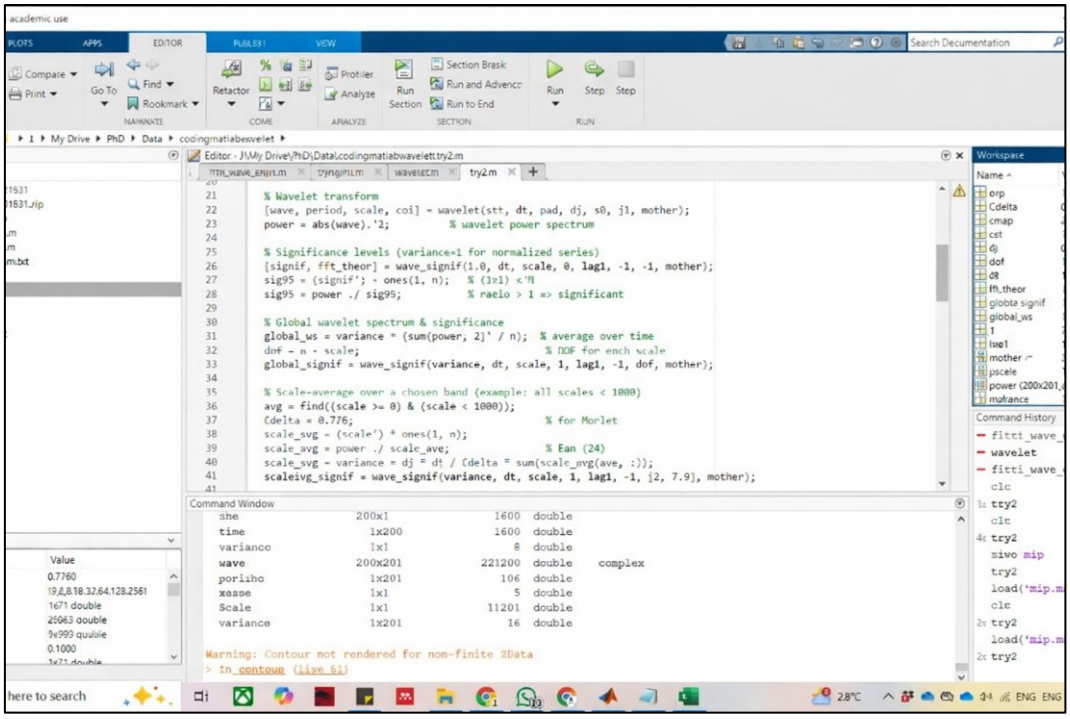


Figure 2. MATLAB interface for data programming.

Implementation of continuous wavelet transform (CWT) for AFR signal analysis, showing the resulting wavelet power spectrum (WPS) and the derived global wavelet spectrum (GWS) used to characterize transient AFR variations in the time–frequency domain.

2.3 Limitations of the Work

The study did not explicitly define controlled acceleration and deceleration profiles, as the primary objective was to capture naturalistic driving behavior. Most of the investigation focused on reflecting real-world driving styles under typical urban and suburban conditions.

Consequently, no emissions data were measured, since the study relied on on-board diagnostics (OBD) data collection rather than controlled engine or emissions testing.

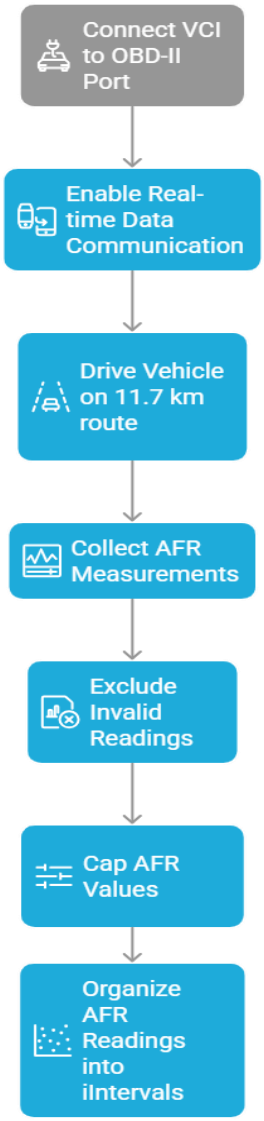


Figure 3. Flowchart for vehicle recording and analysis

3. Results and Discussion

Figure 4(a) presents the AFR profile for aggressive driving, while Figure 5(b) illustrates the trend for moderate driving. The comparison reveals notable differences in AFR stability between the two conditions. During aggressive driving, the AFR fluctuated widely, ranging between 13.0 and 18.0. The signal shows rapid oscillations with frequent excursions towards lean conditions (above 16.5) and occasional drops to richer values near 13.0. This irregular pattern reflects the influence of sudden throttle inputs and abrupt load variations, which demand continuous fuel delivery corrections by the engine management system. Such fluctuations indicate that the combustion process is less stable under aggressive conditions. In contrast, moderate driving produced a narrower AFR distribution, typically ranging between 14.0 and 17.5. While fluctuations remained, they occurred within a more consistent band, with fewer extreme deviations. This relative stability suggests that smoother acceleration and deceleration kept the AFR closer to the stoichiometric value (≈14.7), thereby facilitating more

balanced combustion and improving catalytic converter efficiency. The implications of these variations are significant.

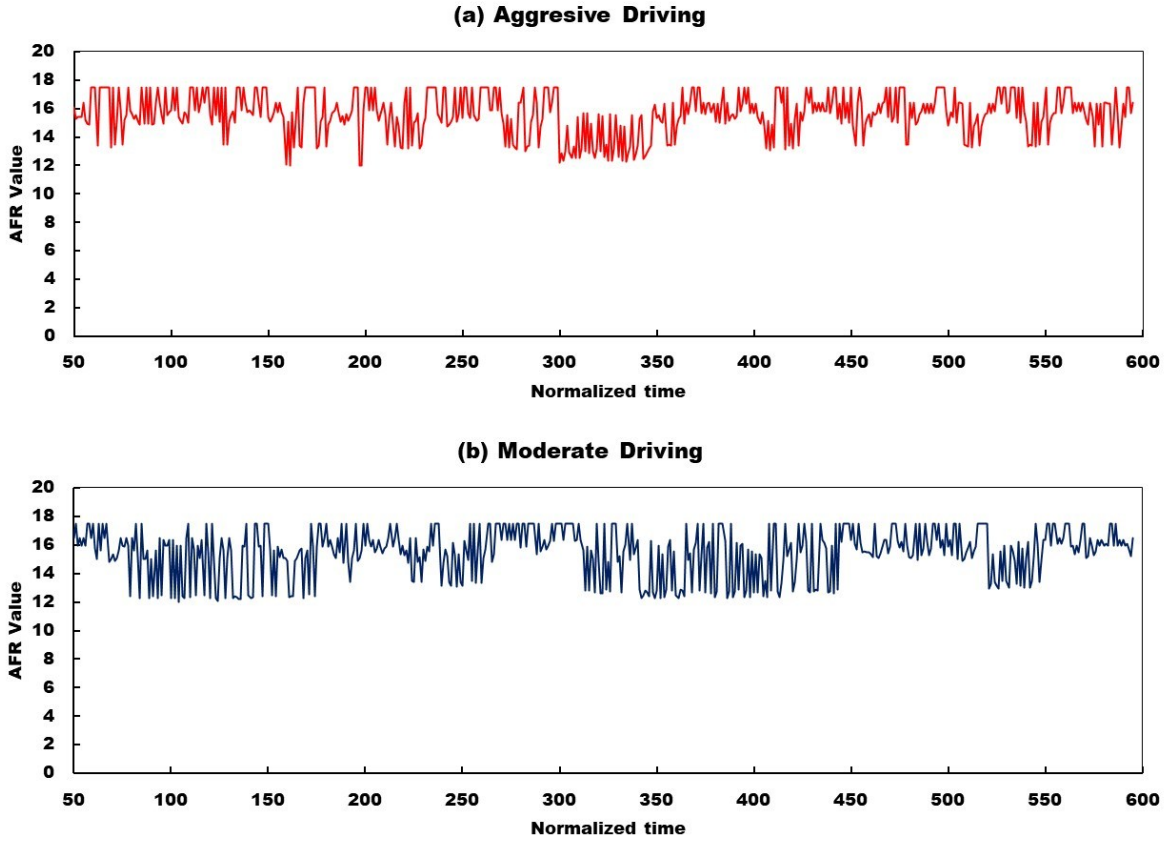


Figure 4. Comparison of AFR dynamics under (a) aggressive and (b) moderate driving conditions

The broader AFR swing observed during aggressive driving may increase transient fuel consumption and adversely affect emissions. Rich excursions (≈13.0–13.5) are associated with incomplete combustion, leading to higher hydrocarbon (HC) and carbon monoxide (CO) emissions. Conversely, lean excursions (≈17.0–18.0) may promote elevated nitrogen oxide (NO_x) formation [26]. On the other hand, the narrower AFR band in moderate driving (≈14.0–17.5) indicates a more controlled combustion process, which supports both fuel economy and emission reduction. Overall, the findings highlight the influence of driving style on AFR dynamics. Aggressive behavior amplifies AFR variability, whereas moderate driving maintains the mixture within a range more favorable for efficient combustion and lower emissions. These results reinforce the importance of adopting smoother driving practices to improve energy efficiency and minimize environmental impact [27].

Figure 5 presents a boxplot analysis of aggressive and moderate driving styles. In this analysis, the mean, median, and outlier values for each driving style can be clearly observed. This is valuable because the boxplot effectively illustrates the characteristics of different driving behaviors and their potential impact on exhaust emissions and fuel consumption. Boxplots provide a concise way to assess whether a driving style reduces emissions, improves fuel efficiency, or increases exhaust output. Unlike conventional graphs, which require multiple figures to show the maximum, minimum, mean, median, and outliers, a single boxplot captures all these statistical features in a single visualization. This makes it a powerful tool for comparing and interpreting the effects of driving behavior on vehicle performance and environmental impact. For aggressive driving, the AFR values ranged from a maximum of 17.5

to a minimum outlier at 12.5. The median was 15.6, with the first quartile (Q1) at 14.9 and the third quartile (Q3) at 16.3. This indicates that 50% of the AFR values were concentrated between 14.9 and 16.3. Despite this relatively narrow interquartile range (IQR = 1.4), the presence of extreme outliers as low as 12.5 highlights the instability in AFR regulation under aggressive throttle changes. Such excursions into richer regions are consistent with transient fuel enrichment during sudden accelerations, which is often employed to avoid knock and maintain power output. In comparison, moderate driving also showed an AFR maximum of 17.5, with outliers observed at 12.25. The median AFR was slightly lower at 15.75, while Q1 and Q3 were 14.8 and 16.3, respectively. This suggests that AFR stability during moderate driving was comparable in terms of spread, but the data distribution was more balanced around the stoichiometric value (≈14.7). Importantly, the lower frequency of extreme outliers indicates that AFR control under moderate driving conditions is less perturbed by sudden fuel-enrichment events. The statistical comparison shows that, while both driving styles operate within similar AFR boundaries (≈12.25–17.5), aggressive driving results in a higher incidence of extreme-rich excursions. These events are associated with incomplete combustion, leading to elevated hydrocarbon (HC) and carbon monoxide (CO) emissions. In contrast, the more balanced AFR distribution observed in moderate driving conditions suggests improved combustion stability, contributing to better fuel economy and reduced pollutant formation. Overall, the boxplot analysis reinforces earlier time-series findings: aggressive driving amplifies AFR variability and increases the likelihood of rich excursions, whereas moderate driving keeps AFR closer to stoichiometric balance, with fewer extreme deviations. This confirms the critical role of driving style in shaping engine combustion behavior and its environmental impact [28].

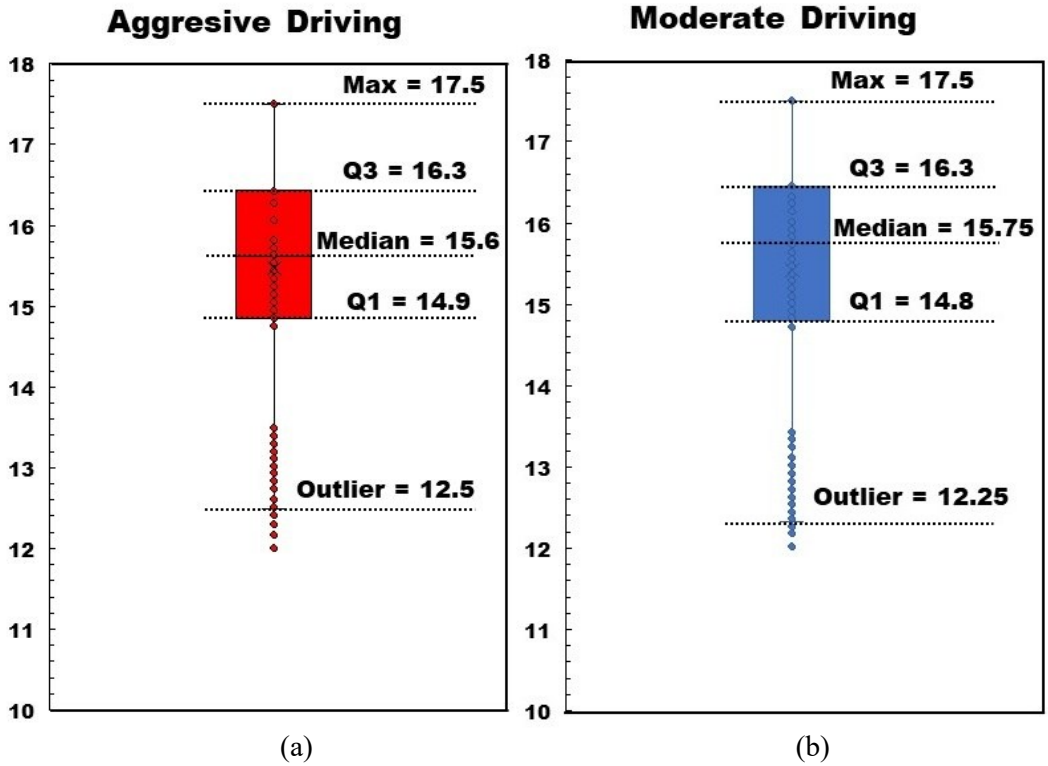


Figure 5. Comparison of key statistical features (max, min, median, quartiles, outliers) for (a) aggressive vs. (b) moderate driving

https://doi.org/10.31436/iiumej.v27i2.4062

Figure 6 shows the AFR frequency distribution for aggressive and moderate driving. The results highlight how different driving behaviors influence the proportion of time the engine operates within specific AFR ranges. Under moderate driving, AFR values are more frequently concentrated in the leaner range of 15.3–16.4, with a relative frequency of approximately 36%, compared to 30% for aggressive driving. This indicates that smoother throttle application allows the engine to operate more consistently in a lean-to-stoichiometric region, which is generally favorable for fuel economy. Moderate driving also shows a higher proportion of values in the 12.0–13.1 range (~18%), reflecting occasional rich mixtures, likely linked to acceleration or load demands. Conversely, aggressive driving demonstrates a more even spread across AFR ranges. While still showing a dominant proportion in the 15.3–16.4 and 16.4–17.5 ranges (~30–31%), aggressive driving exhibits a noticeably higher frequency in the 13.1–14.2 and 14.2–15.3 ranges (12–17%) compared to moderate driving (5–14%). This suggests that aggressive throttle actions lead to more frequent transitions into richer operating zones, corresponding to the increased fuel delivery required during sudden load changes. Taken together, these findings suggest that while both driving modes spend significant time in the 15.3–17.5 range, moderate driving favors leaner AFR stability, whereas aggressive driving is associated with greater variability and richer excursions. From a practical perspective, this implies that moderate driving supports fuel-efficient operation, while aggressive driving, with its richer transients, could contribute to higher fuel consumption and elevated hydrocarbon (HC) and carbon monoxide (CO) emissions.

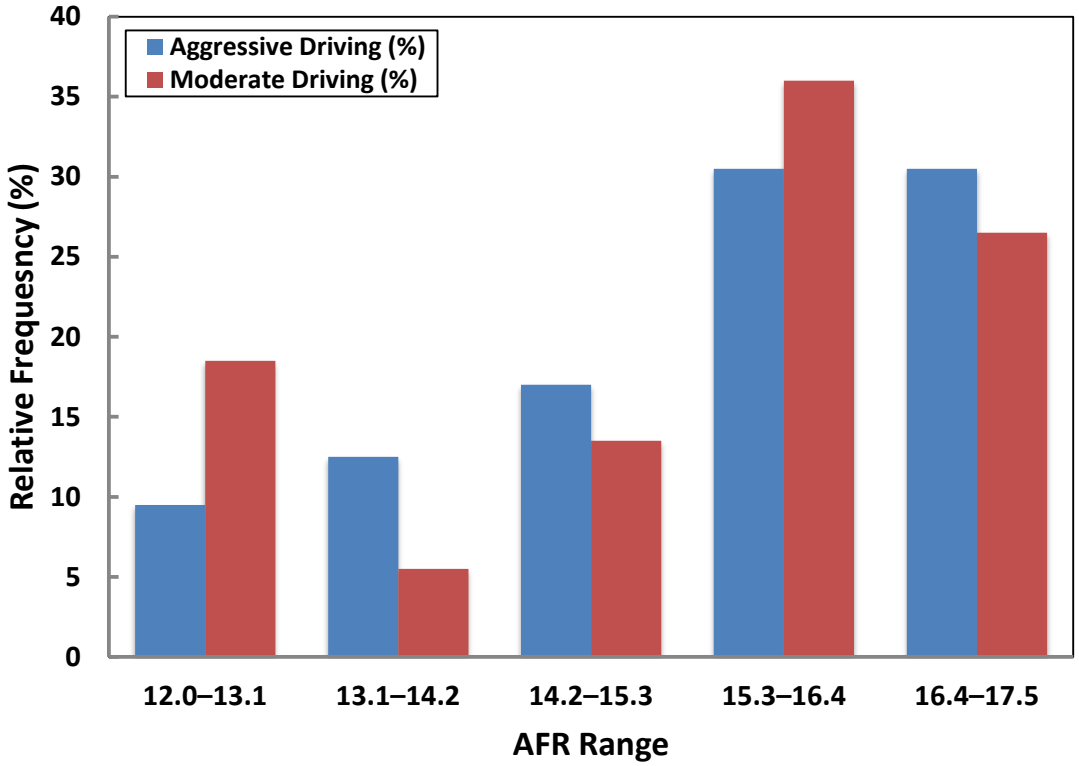
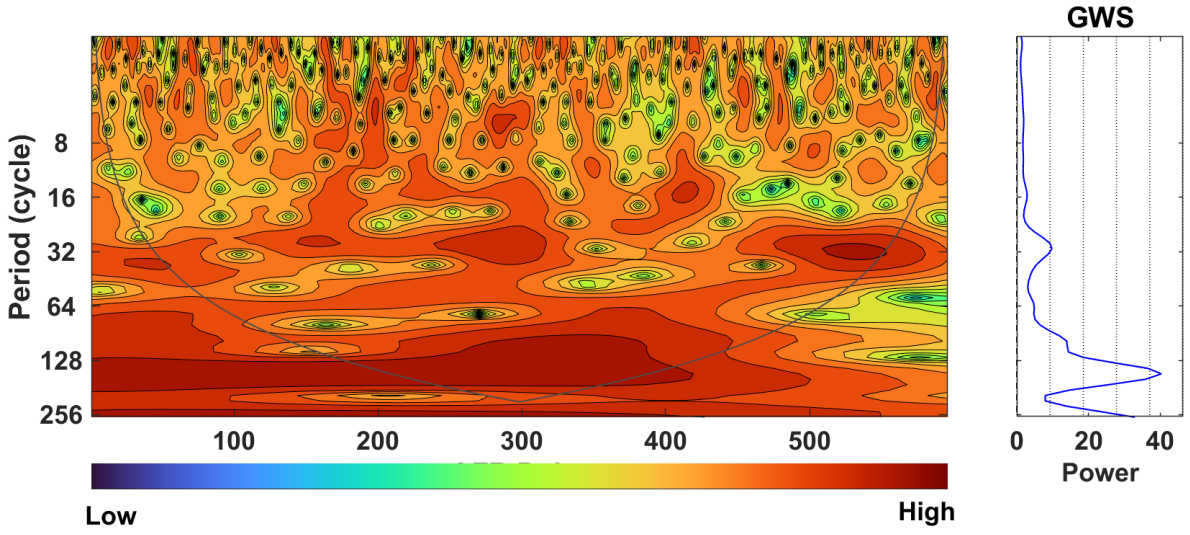


Figure 6. Frequency distribution of Air–Fuel Ratio (AFR) under aggressive and moderate driving

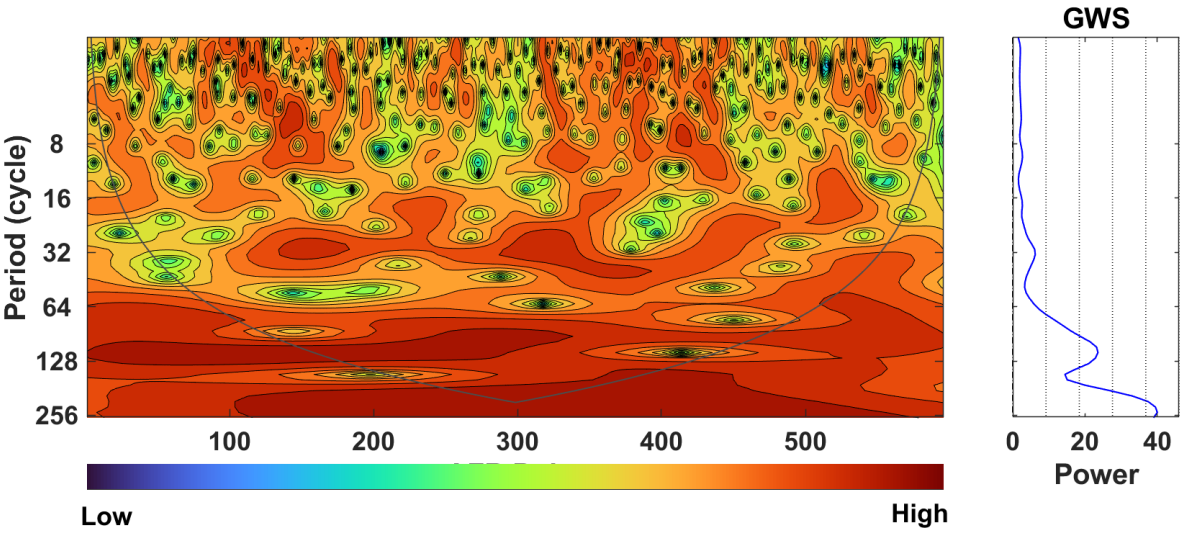
Figure 7(a) illustrates the aggressive driving style, while Figure 7 (b) represents the moderate driving style. These results, derived from the raw dataset, highlight the wavelet power spectrum's (WPS) ability to detect abnormalities in driving behavior. Driving style is widely recognized as a key factor that directly influences fuel efficiency, vehicle performance, and exhaust emissions. By applying both the wavelet power spectrum and the global wavelet spectrum (GWS), it is possible to evaluate how variations in the air–fuel ratio (AFR) reflect

https://doi.org/10.31436/iiumej.v27i2.4062

engine dynamics across different time and frequency domains. The comparative assessment of aggressive and moderate driving reveals a clear distinction in energy distribution, as observed in power levels and cycle ranges. These characteristics are strong indicators of the driver's behavioral condition. For aggressive driving, the wavelet power spectrum reveals a pronounced concentration of high-energy components, quantitatively reflected by dominant global wavelet spectrum peaks reaching approximately 40 power units. These high-energy components are distributed across both low-cycle bands (8–16 cycles) and higher-cycle bands (64–128 cycles), indicating AFR disturbances occurring over multiple time scales due to rapid throttle modulation and braking events.



(a) Aggressive driving style



(b) Moderate driving

Figure 7. Wavelet Power Spectrum (WPS) and Global Wavelet Spectrum (GWS) analysis of (a) aggressive and (b) moderate driving styles

The global wavelet spectrum further confirms these findings, showing sudden peaks reaching up to 40 power units. Multiple dominant peaks are visible across both the lower-cycle region (8–16 cycles) and the higher-cycle region (64–128 cycles), indicating an unstable AFR condition. Such instability in AFR control leads to poor fuel-consumption management, elevated exhaust emissions, and greater mechanical stress on the vehicle's powertrain. In

contrast, moderate driving exhibits a lower overall spectral energy density, with global wavelet spectrum power values predominantly between 10 and 20 units and a broad dominant peak centered at approximately 20 units. This pattern corresponds to power values ranging between 10 and 20 units in the global wavelet spectrum (GWS). The energy distribution is concentrated mainly within the 32–128 cycle range, reflecting smoother engine operation, steadier driving conditions, and improved throttle control. The peak for moderate driving is observed at approximately 20 units, where it appears broad and sustained, particularly across the 32–128 cycle range. This indicates more stable combustion dynamics compared to aggressive driving. Isolated smaller peaks are also present within the 128–256 cycle range, but these remain less significant and do not alter the overall stability of the spectrum. Although some peaks remain slightly elevated, the overall trend in moderate driving shows a steadier, more consistent increase in energy compared to aggressive driving. In particular, the dominant global wavelet spectrum under moderate behavior is concentrated at longer cycles, unlike aggressive driving, which displays multiple scattered peaks. From an engineering perspective, this driving style enhances engine durability, improves driving comfort, and supports eco-driving efficiency. The more stable AFR control under moderate driving contributes to better fuel consumption management and lower vehicle emissions [29]. By comparison, aggressive driving produces a GWS peak of up to 40 units, nearly double the value of moderate driving, highlighting its greater instability and inefficiency.

4. CONCLUSIONS

Aggressive and moderate driving behaviors produce clearly distinguishable air–fuel ratio (AFR) patterns, with aggressive driving causing greater AFR instability and moderate driving supporting more consistent AFR regulation. The results show that aggressive driving, characterized by frequent and abrupt throttle inputs, leads to higher AFR variability, repeated rich excursions with AFR values reaching as low as 12.5, and stronger oscillatory behavior, as reflected by global wavelet spectrum peak magnitudes of up to 40 units across multiple cycle ranges. In addition, rich AFR operating regions occur more frequently under aggressive driving, accounting for approximately 12–17% of total driving time, indicating repeated transient fuel enrichment and greater deviation from the stoichiometric condition. In contrast, moderate driving maintains AFR values closer to the stoichiometric ratio of approximately 14.7, demonstrates a higher proportion of smoother lean operation within the 15.3–16.4 AFR range at about 36%, and produces fewer extreme AFR outliers in the boxplot analysis, confirming lower AFR dispersion and more stable combustion control. Overall, the findings demonstrate that driving behavior has a direct, measurable effect on AFR stability, distribution, and transient enrichment patterns, with aggressive driving contributing to less stable AFR regulation and moderate driving promoting smoother, more efficient engine operation. Future work should extend this analysis to different vehicle types, engine technologies, road gradients, and traffic conditions to establish a more generalizable relationship between driving behavior and AFR dynamics.

ACKNOWLEDGEMENT

The authors would like to thank Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA) for providing financial support under the research grant scheme No. PGRS230386 and laboratory facilities.

REFERENCES

[1] A. Boretti, "Advancing sustainable mobility: Integrating flywheel kinetic energy recovery systems with high-efficiency hydrogen internal combustion engines," *Int. J. Hydrogen Energy*, vol. 125, pp. 354–363, 2025.

[2] N. Venkatachalam, A. P. V. Manoharan, K. Ravichandran, P. V. Aravinth, and T. H. Sardar, "Speed Breaker and Vehicle Accident Detection with Alert Sensors," in *Applications of Computational Learning and IoT in Smart Road Transportation System*, Springer, 2025, pp. 17–36.

[3] C. A. MacCarley, "Engines and Fuels," in *Non-Petroleum Automotive Transportation*, Springer, 2025, pp. 61–96.

[4] Y. Ling, Z. Ma, Y. Song, Q. Zhang, X. Weng, and X. Ma, "Bus driver deceleration behavior modeling at intersections using multi-source on-board sensor data," *J. Public Trans.*, vol. 27, p. 100123, 2025.

[5] F. Haque and M. A. Abas, "Review of Driving Behavior Towards Fuel Consumption and Road Safety," *Jurnal Mekanikal*, 2018.

[6] E. G. Giakoumis, *Driving and engine cycles*, vol. 1. Springer, 2017.

[7] T. Wei, T. Zhu, and M. Zhang, "Impact of driving style on rear-end crash risk in leading vehicle emergency braking situations," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, p. 09544070251323635, 2025.

[8] A. Moawad, M. Zebiak, J. Han, D. Karbowski, Y. Zhang, and A. Rousseau, "Effect of adaptive cruise control on fuel consumption in real-world driving conditions," *Nat. Commun.*, vol. 15, no. 1, p. 10016, 2024.

[9] M. S. H. Nahid, T. Z. Tila, and T. S. Seecharan, "Detecting emotional arousal and aggressive driving using neural networks: A pilot study involving young drivers in Duluth," *Sensors*, vol. 24, no. 22, p. 7109, 2024.

[10] M. Kumar and T. Shen, "Estimation and feedback control of air-fuel ratio for gasoline engines," *Control Theory and Technology*, vol. 13, no. 2, pp. 151–159, 2015.

[11] B. Daham, H. Li, G. E. Andrews, J. Tate, K. Ropkins, and M. Bell, "Driver Variability Influences on Real World Emissions at a Road Junction using a PEMS," SAE Technical Paper, 2010.

[12] D. Rimpas, A. Papadakis, and M. Samarakou, "OBD-II sensor diagnostics for monitoring vehicle operation and consumption," *Energy Reports*, vol. 6, pp. 55–63, Feb. 2020, doi: 10.1016/j.egyr.2019.10.018.

[13] A. Zeb *et al.*, "ON-BOARD DIAGNOSTIC (OBD-II) BASED CYBER PHYSICAL SYSTEM FOR ROAD BOTTLENECKS DETECTION," 2022.

[14] G. Padmanaban, F. Feng, E. Dai, A. Saini, G. Hu, and Y. Zhao, "A Comparative Analysis of Acceleration and Deceleration Profiles for Aggressive Driving Styles and Fuel Economy Test Cycles," SAE Technical Paper, 2025.

[15] B. Li *et al.*, "Representing-behavior-based online driving style recognition and its road verification," *Eng. Appl. Artif. Intell.*, vol. 156, p. 111241, 2025.

[16] M. S. Labbo, X. Jiang, S. A. Wada, G. J. de Dieu, and N. D. Bala, "Road rage patterns in Nigeria: A multi-level latent class analysis," *Case Stud. Transp. Policy*, p. 101471, 2025.

[17] K. ŚLUSARCZYK and R. Jurecki, "The use of a vehicle simulator for eco-driving research," *Combustion Engines*, 2025.

[18] L. Zhang, Z. Zhu, Z. Zhang, G. Song, Z. Zhai, and L. Yu, "An improved method for evaluating eco-driving behavior based-on speed-specific vehicle-specific power distributions," *Transp. Res. D Transp. Environ.*, vol. 113, p. 103476, 2022.

[19] A. Adavikottu and N. R. Velaga, "Analysis of speed reductions and crash risk of aggressive drivers during emergent pre-crash scenarios at unsignalized intersections," *Accid. Anal. Prev.*, vol. 187, p. 107088, 2023.

[20] A. Adavikottu, N. R. Velaga, and S. Mishra, "Modelling the effect of aggressive driver behavior on longitudinal performance measures during car-following," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 92, pp. 176–200, 2023.

[21] Z. Ma and Y. Zhang, "Driver-automated vehicle interaction in mixed traffic: Types of interaction and drivers' driving styles," *Hum. Factors*, vol. 66, no. 2, pp. 544–561, 2024.

[22] A. Adavikottu and N. R. Velaga, "Modeling the impact of driving aggression on lane change performance measures: steering compensatory behavior, lane change execution duration and crash probability," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 103, pp. 526–553, 2024.

[23] T. Kerwin and B. J. Bushman, "Measuring the perception of aggression in driving behavior," *Accid. Anal. Prev.*, vol. 145, p. 105709, 2020.

[24] M. R. Saxena and R. K. Maurya, "Characterization of cycle-to-cycle variations in conventional diesel engine using wavelets," in *Advances in internal combustion engine research*, Springer, 2017, pp. 135–155.

[25] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bull. Am. Meteorol. Soc.*, vol. 79, no. 1, pp. 61–78, 1998.

[26] Mashruk *et al.*, "Perspectives on NO X Emissions and Impacts from Ammonia Combustion Processes," *Energy & Fuels*, vol. 38, no. 20, pp. 19253–19292, 2024.

[27] M. Batallas and P. Molina, "Developing a Methodology to Reduce Fuel Consumption and Classify Driving Styles for a Fleet of Vehicles," in *International Conference on Science, Technology and Innovation for Society*, Springer, 2024, pp. 185–194.

[28] H. Abediasl, A. Ansari, V. Hosseini, C. R. Koch, and M. Shahbakhti, "Real-time vehicular fuel consumption estimation using machine learning and on-board diagnostics data," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 238, no. 12, pp. 3779–3793, 2024.

[29] S. Yoo, J. Shin, and S.-H. Choi, "Machine learning vehicle fuel efficiency prediction," *Sci. Rep.*, vol. 15, no. 1, pp. 1–20, 2025.

https://doi.org/10.31436/iiumej.v27i2.4068

Design Optimization and Structural Evaluation of a Hybrid Frame for Next-Generation Electric Scooters

SRIDHAR R[1*], MAHANTESH M MATH[1], SUJAN CHAKRABORTY[1], SOURABHA H[1], VIVEKANAND S GOGI[2], JEEVATHITH R[1]

[1]*Department of Mechanical Engineering, R V College of Engineering, Bangalore, 560059, India*
[2]*Department of Industrial Engineering & Management, R V College of Engineering, Bangalore, 560059, India*

**Corresponding author: sridharr@rvce.edu.in*

(Received: 27 November 2025; Accepted: 2 February 2026; Published online: 10 May 2026)

ABSTRACT: The increasing adoption of electric scooters has introduced unique challenges in chassis design, where packaging constraints imposed by predefined outer shells often conflict with structural efficiency. This study presents the conceptual design, structural analysis, and validation of a hybrid chassis tailored for an electric scooter within such constraints. The frame integrates a step-through central region with a trellis-inspired rear subframe and a reinforced headstock. Three design iterations were modeled in SolidWorks and evaluated using finite element analysis (ANSYS) for static loads (1–3G), fatigue under cyclic loading, equivalent static impact loads, and modal behavior. Analytical stress checks and literature comparisons supported the simulations. Under 3G static loading, the final steel frame exhibited a maximum deformation of 3.16 mm with a minimum factor of safety of 1.48. Fatigue analysis predicted a life of 1.75×10^6 cycles, while modal analysis confirmed the first natural frequency at 28.56 Hz, above the expected excitation frequencies, and thus free from resonance. Although crash simulations showed localized yielding in severe impacts, the frame maintained integrity under operational conditions. These findings demonstrate that high structural performance can be achieved within strictly styling-defined packaging envelopes, validating the hybrid trellis–step-through configuration for electric scooters.

ABSTRAK: Peningkatan penggunaan skuter elektrik telah menyebabkan cabaran unik pada reka bentuk casis, di mana kekurangan pada pembungkusan, yang ditentukan oleh struktur luaran, sering bercanggah dengan kecekapan struktur. Kajian ini membentangkan reka bentuk konseptual, analisis struktur, dan pengesahan bagi casis hibrid yang direka khusus untuk skuter elektrik. Rangka casis ini mengintegrasi bahagian tengah jenis telus langkah dengan subrangka belakang berinspirasi jalaran serta stok kepala yang diperkukuh. Tiga iterasi reka bentuk dimodel menggunakan SolidWorks dan dinilai melalui analisis unsur terhingga menggunakan ANSYS bagi beban statik (1–3G), keletihan di bawah beban berkitar, beban impak statik setara, serta tingkah laku mod. Pengesahan analitikal terhadap tegasan serta perbandingan dengan kajian terdahulu turut menyokong dapatan simulasi. Di bawah beban statik 3G, rangka keluli akhir menunjukkan ubah bentuk maksimum sebanyak 3.16 mm dengan faktor keselamatan minimum 1.48. Analisis keletihan meramalkan jangka hayat sebanyak 1.75×10^6 kitaran, manakala analisis mod mengesahkan frekuensi asli pertama pada 28.56 Hz, iaitu melebihi frekuensi yang diuji dan bebas resonans. Walaupun simulasi perlanggaran menunjukkan luluh setempat di bawah impak teruk, rangka masih kukuh ketika beroperasi. Dapatan ini menunjukkan bahawa prestasi struktur yang tinggi boleh dicapai dalam persekitaran pembungkusan yang ketat berasaskan reka bentuk estetik, sekaligus mengesahkan keberkesanan konfigurasi hibrid jalaran–telus langkah pada skuter elektrik.

https://doi.org/10.31436/iiumej.v27i2.4068

KEYWORDS: *Electric scooter, hybrid chassis, trellis frame, finite element analysis, packaging constraints.*

1. INTRODUCTION

The structural design of two-wheeler chassis has long been explored in the motorcycle domain, where frame types such as double-cradle, backbone, perimeter, monocoque, and trellis layouts have undergone extensive optimization. Among these, the trellis frame has emerged as particularly effective, offering high torsional rigidity with relatively low material use due to its triangulated geometry [1]. In parallel, suspension design plays a critical role in handling and load transfer, and foundational studies on chassis–suspension interaction have provided a strong basis for modern two-wheeler development [2]. In contrast, scooter chassis design has historically been dominated by step-through and underbone configurations, which prioritize ease of use and compactness. While effective for urban commuting, these layouts inherently feature interrupted load paths that reduce stiffness and fatigue resistance compared to motorcycle frames [3]. The removal of the engine block further eliminates a key contributor to stiffness, while the introduction of bulky batteries and hub motors alters weight distribution and packaging constraints, complicating the design process. Computational design and stress analysis methods have shown potential in identifying safe stress limits and optimizing weight distribution for electric motorcycle frames [4]. Similar approaches applied to electric bike chassis have demonstrated how finite element simulations improve durability while supporting lightweighting goals [5]. Together, these studies highlight the growing importance of simulation in guiding structural design decisions for two-wheelers.

The transition to electric scooters introduces additional structural challenges. Manufacturers often impose styling envelopes that constrain frame geometries, forcing designers to balance accessibility with rigidity. To address this, researchers have proposed hybrid frame concepts that integrate scooter ergonomics with trellis-inspired geometries. Shivhare et al. [6] modeled a compact, lightweight scooter using finite element analysis, highlighting how triangulated layouts reduce weight without sacrificing stiffness. Goh and Chun [7] further advanced this idea by applying trellis-frame principles to electric scooters, demonstrating measurable improvements in both structural efficiency and vibration safety. Virtual evaluation methods now play a central role in validating chassis configurations. Finite Element Analysis (FEA) enables simulation of static loads, fatigue conditions, impact scenarios, and modal responses, offering a comprehensive framework for performance evaluation. Karthik [8] emphasized the value of simplified fabrication combined with reinforced member placement, achieving safety margins above 2.0 without excessive material use. Ajhay Babu [9] conducted a computational study of an electric two-wheeler frame, highlighting the importance of assessing natural frequencies and validating crashworthiness across multiple load cases.

Material optimization has also emerged as a decisive factor in electric vehicle chassis design. Charkha et al. [10] conducted a comparative analysis of materials for a three-wheeled electric vehicle, concluding that while composites offer lightweighting, structural steels provide higher fatigue safety margins under urban use. Rahmadi and Mursalin [11] analyzed a chopper-style motorcycle frame, reinforcing the importance of member geometry in achieving high durability and improved rider safety. Beyond conventional motorcycles, several studies have focused on small-format electric two-wheelers. Srinu et al. [12] applied computational analysis to an electric bike chassis, demonstrating how load-bearing components can be optimized for stiffness while reducing material consumption. Teja et al. [13] developed an

https://doi.org/10.31436/iiumej.v27i2.4068

electric moped chassis, emphasizing packaging efficiency and component integration within tight geometric envelopes. Vashist et al. [14] investigated electric two-wheeler frames under variable loading conditions, confirming the critical role of frame geometry in determining structural reliability. Biswal et al. [15] introduced an innovative sliding-frame electric bike design, showcasing adaptability and improved load management compared to fixed-frame configurations.

Taken together, these studies emphasize the need to integrate material selection, structural layout, fatigue validation, modal assessment, and manufacturability into a cohesive framework for electric scooter design. Despite these advances, hybrid trellis–step-through architectures remain underexplored. The present work addresses this gap by developing and analyzing such a chassis through virtual structural evaluation, aiming to balance accessibility, safety, and performance in urban mobility.

2. MATERIALS AND METHODOLOGY

2.1. Requirement Analysis

The chassis was designed to meet the performance needs of an urban step-through electric scooter while staying within the predefined packaging and ergonomic limits. Key requirements included a maximum payload capacity of 150 kg (rider, pillion, and luggage) plus an additional 30 kg for the motor and battery assembly. For Indian road conditions, the scooter required a minimum ground clearance of 150 mm and a wheelbase of about 1290 mm. The primary use case was daily urban commuting, in which the frame would be subjected to repeated loading from speed breakers, potholes, and surface irregularities. These parameters formed the basis for geometry design, material selection, and structural validation.

2.2. Body Shell Constraint Modeling

A 3D CAD model of the scooter's body shell was created in Autodesk Alias to serve as the primary packaging envelope for the chassis. The packaging model, shown in Figure 1a, defined the maximum permissible volume for structural members while preserving the step-through clearance essential for rider ergonomics. Mounting zones for key subsystems, such as the battery, motor, and suspension, were incorporated to facilitate integration.



Figure 1. (a) Chassis Packaging Model (b) Outer Body Shell

The outer body shell, shown in Figure 1b, imposed strict spatial limitations, particularly around the headstock, footboard, and seat regions, disrupting conventional load paths and restricting standard triangulated layouts. Consequently, the chassis geometry was developed to

fit within these allowable regions, balancing packaging constraints and structural efficiency. Dimensional limits included a total length of approximately 1900 mm, a wheelbase of roughly 1290 mm, and a minimum ground clearance of 185 mm.

2.3. Geometrical Modeling of Chassis

The chassis was developed using the CAD software (SolidWorks) in a packaging-driven approach, with a fixed body shell geometry serving as the governing design envelope. The objective was to integrate a structurally efficient hybrid step-through and trellis frame within the constrained volume while maintaining ergonomics, manufacturability, and serviceability.

The frame was divided into three functional zones. The front headstock region comprised a curved loop tube reinforced with gusseted members to carry steering and impact loads. The central spine and floorboard zone supported the battery and rider loads through larger-diameter 26.9 mm tubes. The rear subframe and swingarm mount were designed as a triangulated trellis structure to resist propulsion and suspension forces while providing impact absorption.

Table 1. Tube Profiles of Chassis

Sl. No	Tube size	Application zone
1	Ø33.7 mm × 2 mm	Headstock and frontal loop members
2	Ø26.9 mm × 2 mm	Main spine and floor support tubes
3	Ø21.3 mm × 2 mm	Secondary supports and rear triangulation

Circular steel tubes of Ø33.7 mm, Ø26.9 mm, and Ø21.3 mm with 2 mm wall thicknesses listed in Table 1 were employed. This configuration balanced stiffness and weight, resulting in an approximately 20 kg total assembly including brackets and welds. For preliminary member sizing validation, a 3G design case was considered using the scooter's gross design mass of 250 kg, including the rider, frame, and battery. This corresponds to a total vertical force given by Eq. (1).

$$F = nmg \tag{1}$$

where $n = 3$ is the load factor for the 3G design case, $m = 250$ kg, and $g = 9.81$ m/s^2. Thus, $F = 7357.5$ N. With an assumed moment arm of L=150 mm, the bending moment (M) was computed using Eq. (2)

$$M = FL \tag{2}$$

With $M = 1103.6$ Nm, for a Ø26.9 mm tube with 2 mm wall thickness, the inner diameter is $d = D - 2t = 22.9$ mm. The second moment of area, I, for a hollow circular tube, is given by $I = \frac{\pi}{64}(D^4 - d^4)$. The corresponding section modulus (S) is calculated using Eq. (3)

$$S = \frac{I}{D/2} = \frac{\pi}{32}\left(\frac{D^4 - d^4}{D}\right) \tag{3}$$

For $D = 26.9$ mm and $d = 22.0$ mm, then $S = 907.3$ mm^3. The bending stress (σ) was then calculated using Eq. (4)

$$\sigma = \frac{M}{S} \tag{4}$$

Using $M = 1103.6\ N \cdot m = 1.1036 \times 10^6 N.mm$, $\sigma = \frac{1.1046 \times 10^6}{907.33} = 1216.4$ MPa.

This bending stress exceeds the 350 MPa yield strength of E350 steel if the Ø26.9 mm tube is assumed to act as a single isolated member under the full 3G load. However, in the actual frame, the load is distributed among multiple interconnected members and further

supported by welded joints and triangulated structural paths. Therefore, this simplified calculation represents a conservative single-member validation case rather than the actual stress state of the complete frame.

Three CAD iterations were assessed to optimize structural performance. Iteration 1 employed a minimal bracing layout with a single central spine, but offered limited rigidity. Iteration 2 introduced frontal reinforcement and partial rear bracing, which improved stiffness yet still lacked mid-span support. Iteration 3, the final design, featured a fully triangulated rear trellis, multiple frontal braces, and gusset reinforcements, delivering the best balance of stiffness, weight, and manufacturability, and was therefore chosen for further analysis.

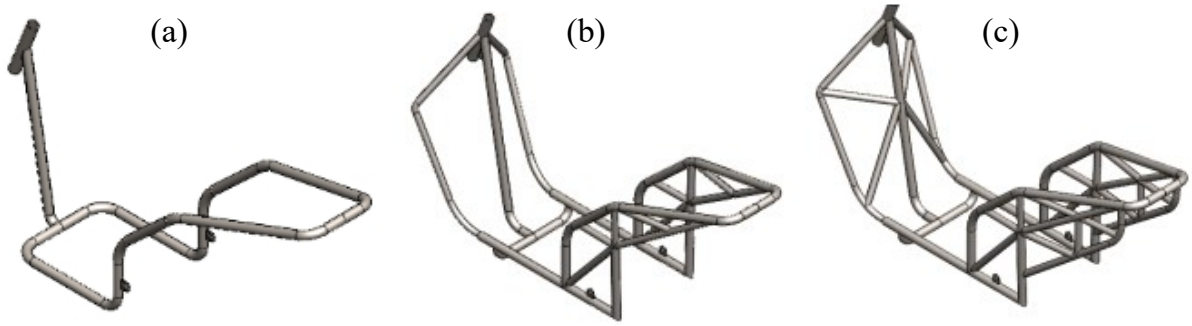


Figure 2. (a) 1st Chassis Iteration, (b) 2nd Chassis Iteration (c) 3rd Chassis Iteration

The final line sketch in Figure 3(a) confirmed key handling parameters: headstock angle 21.5°, wheelbase 1282 mm, trail 90.6 mm, and the center of gravity ~437 mm above ground. Footboard clearance was 232 mm, sufficient for Indian urban road conditions. These values were carried over into the CAD model (Figure 3(b)) for subsequent simulation.

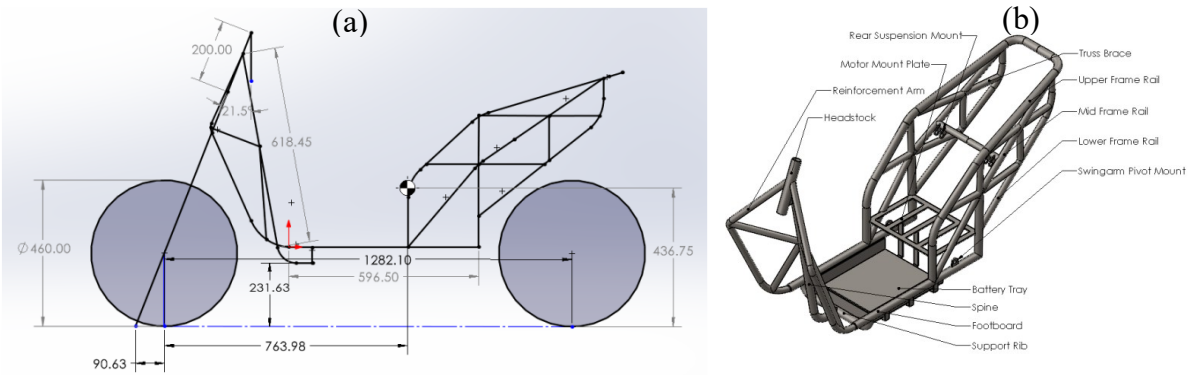


Figure 3. (a) Final Chassis Geometry Line Sketch (b) Final Chassis Assembly

A non-functional scaled 3D-printed model was fabricated using fused deposition modeling (FDM) for design communication and ergonomic validation (Figure 4). The chassis was printed in PLA to represent structural members, while the body shell was printed in a contrasting color to illustrate packaging compliance.

https://doi.org/10.31436/iiumej.v27i2.4068



Figure 4. Scaled 3D-printed models of (a) hybrid trellis chassis and (b) complete scooter with a body shell.

2.4. Material Selection

The choice of frame material plays a critical role in ensuring structural stiffness, fatigue resistance, manufacturability, and cost-effectiveness. Since the chassis must withstand both static and dynamic urban loads, a performance-based selection methodology was employed using Ashby material indices (Figure 5a). The key requirements for selection included stiffness to minimize deformation under operational loads, strength to prevent yielding and fatigue failure, manufacturability in terms of weldability, tube availability, and ease of fabrication, and cost-effectiveness for large-scale production. Based on these criteria, three candidate materials were shortlisted: AISI 4130 Chromoly steel, which offers high strength and excellent fatigue resistance; IS 2062 E350 structural steel, which provides moderate strength, good availability, and is highly economical; and Aluminum 6061-T6, which is lightweight, corrosion-resistant, and already widely used in electric vehicle chassis.

The stiffness performance index was derived from the cantilever deflection relationship and is given by Eq. (5).

$$Stiffness_PI = \frac{\sqrt{E}}{\rho} \tag{5}$$

where $Stiffness_PI$ is the stiffness performance index, E is Young's Modulus, and ρ is the density of the material.

Since strength also plays a decisive role in frame integrity, a composite performance index was introduced to account for both stiffness and strength while incorporating mass sensitivity. This relationship is given by Eq. (6):

$$Composite_PI = \frac{\sigma_y \sqrt{E}}{\rho} \tag{6}$$

where $Composite_PI$ is the composite performance index and σ_y is the yield strength.

Table 2. Mechanical properties and performance indices of candidate materials

SL No	Property	IS 2062 E350	AISI 4130 Chromoly	Aluminum 6061-T6
1	Young's Modulus (GPa)	200	210	68.9
2	Yield Strength (MPa)	350	560	276
3	Density (kg/m³)	7850	7850	2700
4	√E/ρ (Stiffness PI)	1.80	1.84	3.05
5	Composite Index*	796.8	1271.3	384.8

**Higher values indicate a better stiffness–strength–mass balance.*

https://doi.org/10.31436/iiumej.v27i2.4068

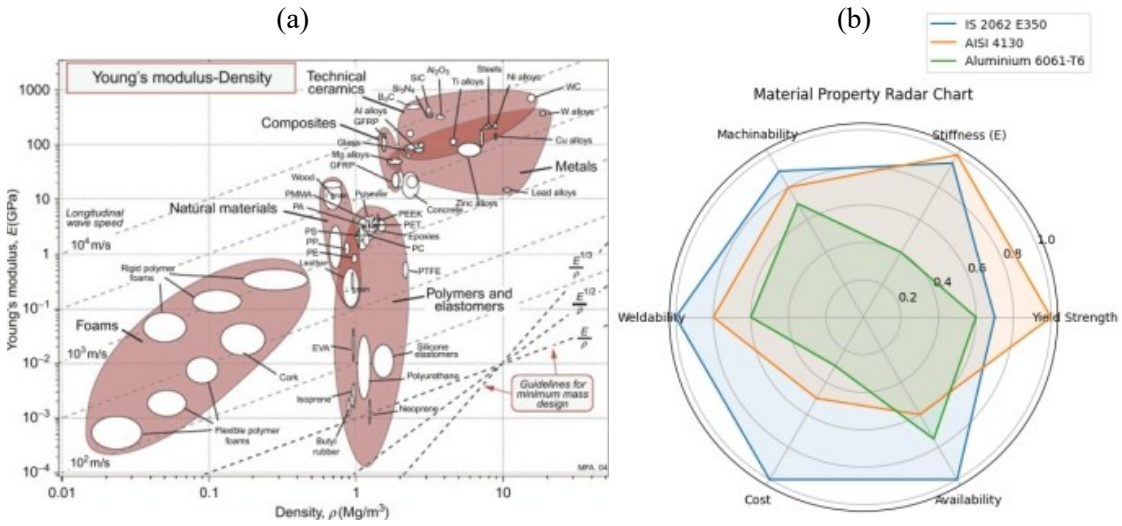


Figure 5. (a) Young's modulus vs. density (Ashby, 2005) (b) Composite Performance Index Radar Chart

The comparative results of these indices are presented in Table 2 and Figure 5b. The evaluation highlights that AISI 4130 exhibits the highest overall performance owing to its superior combination of stiffness and strength. Aluminum 6061-T6 provides an excellent stiffness-to-weight ratio, but its lower modulus limits its resistance to deformation under load. IS 2062 E350 strikes a balanced trade-off among strength, cost, and availability, making it an attractive option for cost-sensitive, large-scale production. While AISI 4130 is technically the most capable material, IS 2062 E350 emerges as a practical choice, particularly in the context of mass manufacturing, where cost and supply-chain accessibility are decisive.

2.5. Finite Element Analysis

Finite Element Analysis (FEA) was carried out in ANSYS Workbench to evaluate the structural performance of the hybrid step-through scooter chassis. Four simulation domains were investigated: static structural, fatigue, impact, and modal analyses to evaluate stiffness, strength, durability, and vibration safety. The chassis was modeled in IS 2062 E350 structural steel, with Aluminum 6061-T6 used as a benchmark for comparative evaluation. The CAD model from SolidWorks was simplified prior to analysis by excluding non-structural parts, such as brackets and fasteners, to reduce computational time.

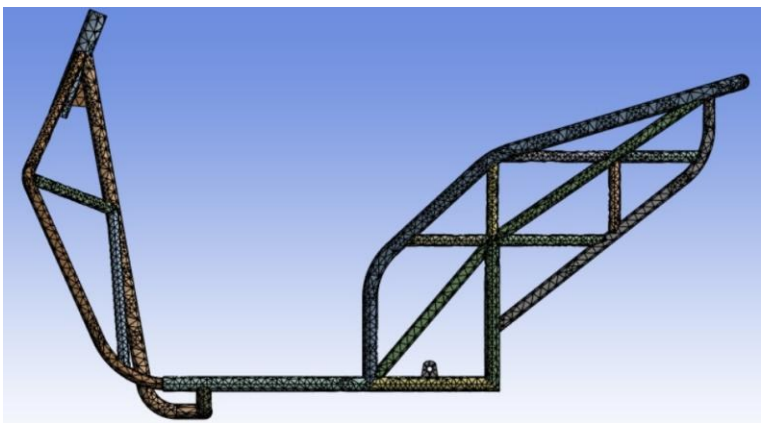


Figure 6. Chassis frame after meshing.

https://doi.org/10.31436/iiumej.v27i2.4068

The frame was meshed using a patch-conforming tetrahedral method. A global element size of 20 mm was adopted, with local refinements of 15 mm applied to four key structural members and a gusset region to better capture local stress concentrations. Further refinement was attempted, but the solver failed to compute beyond this level due to computational resource limits. The final mesh contained approximately 119,161 nodes and 58,220 elements, as shown in Figure 6. This meshing strategy represents a balance between accuracy and computational feasibility.

The boundary conditions for the structural simulation were defined to replicate realistic loading and support scenarios. A fixed support was applied at the rear swingarm mounts, thereby constraining all degrees of freedom at this location. At the headstock, a cylindrical support was introduced to permit both axial and tangential rotation, thereby reproducing the chassis' steering behavior under operating conditions. To evaluate the influence of material properties on overall performance, two material configurations were considered independently in the simulation. The first configuration employed IS 2062 structural steel, while the second utilized Aluminum 6061-T6. This approach enabled a direct comparison of stiffness and stress distribution for the two materials under identical boundary and loading conditions. To maintain conciseness, graphical results are illustrated for steel, while aluminum outcomes are summarized in tables for direct comparison.

2.5.1. Static Structural Analysis

Static analysis quantified the stress distribution and deflection under rider and component loads. Vertical loads corresponding to 1G, 2G, and 3G conditions were calculated using Eq. (1), where m = 100 kg (rider + frame + battery). The resulting loads were distributed across the seat, floorboard, and headstock (Table 3).

Table 3. Load distribution across structural zones

SL No	Zone	Percentage	Load at 1G (N)	Load at 2G (N)	Load at 3G (N)
1	Seat	60%	588.6	1177.2	1765.8
2	Floorboard	25%	245.25	490.5	735.75
3	Headstock	15%	147.15	294.3	441.45

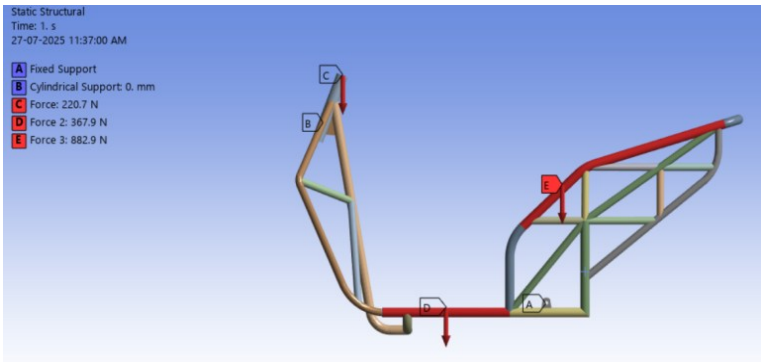


Figure 7. Loads applied to the chassis frame for static analysis.

2.5.2. Fatigue Analysis

Fatigue analysis assessed the long-term durability of the chassis under cyclic loading caused by rider movement, braking, and road-surface irregularities. A cyclic load of 1.5G was applied using the same constraints as in the static analysis. Fatigue life was estimated using the S–N curve approach with Goodman's relation given in Eq. (7)

https://doi.org/10.31436/iiumej.v27i2.4068

$$\sigma_a = \sigma_f'(2N_f)^b \quad (7)$$

Since published S–N data for IS 2062 and 6061-T6 were limited, equivalent literature curves (ASTM A36 structural steel, AA6061-T6 aluminum) were adopted. Missing high-cycle values were extrapolated to 10^8 cycles using standard fatigue trends. For IS 2062 steel, an endurance limit of 160 MPa was used, while aluminum 6061-T6 was evaluated at 95 MPa at 10^7 cycles.

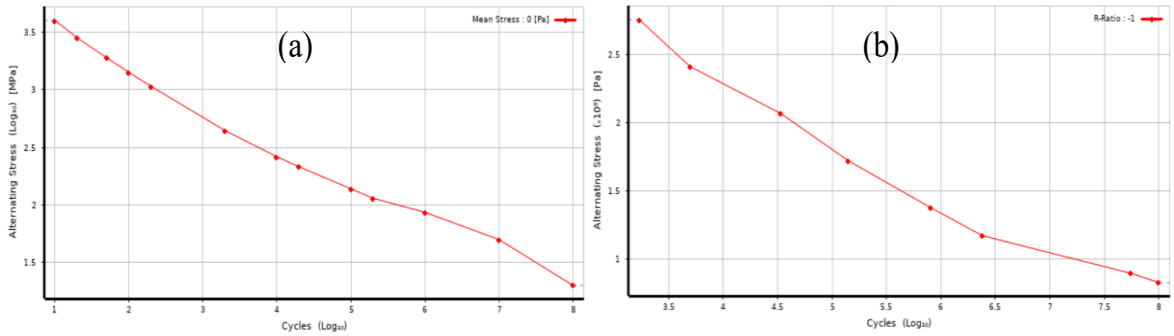


Figure 8. S–N curve for (a) IS 2062 E350 steel and (b) aluminum 6061-T6 alloy

The S–N curves used in fatigue estimation were defined in ANSYS Engineering Data (Figure 8). For IS 2062 E350, alternating stress amplitudes decrease from the order of gigapascals at very low cycles (10–100) to ~50 MPa at 10^7 cycles. For 6061-T6 aluminum, fatigue strength ranged from ~2.75×10^8 Pa at 1.7×10^3 cycles to ~8.27×10^7 Pa at 10^8 cycles, consistent with MIL-HDBK-5H data.

2.5.3. Impact Analysis

Impact loading was modeled using an equivalent static force derived from the work-energy principle. The kinetic energy of the rider–vehicle system was assumed to be absorbed over a prescribed stopping distance. The relationship is given by Eq. (8):

$$F_{eq} = \frac{mv^2}{2d} \quad (8)$$

where m is the effective impact mass, v is the impact velocity, F_{eq} is the equivalent of static impact force, and d is the assumed stopping distance or crush distance. The equivalent impact force ($F_{eq} \approx 1715$ N) was derived from energy conservation using a 100 kg system (rider + vehicle) at 30 km/h (8.33 m/s) and an assumed stopping distance of 0.2 m (conservative for vehicle crush). For conservative localized structural reaction and to match common scooter impact studies, we applied an equivalent static load scaled to represent distributed energy absorption; using an effective force-time conversion and prior literature scaling, the adopted equivalent point load was 1715 N. This approach provides a repeatable, conservative comparative metric while acknowledging limitations versus explicit dynamics.

https://doi.org/10.31436/iiumej.v27i2.4068

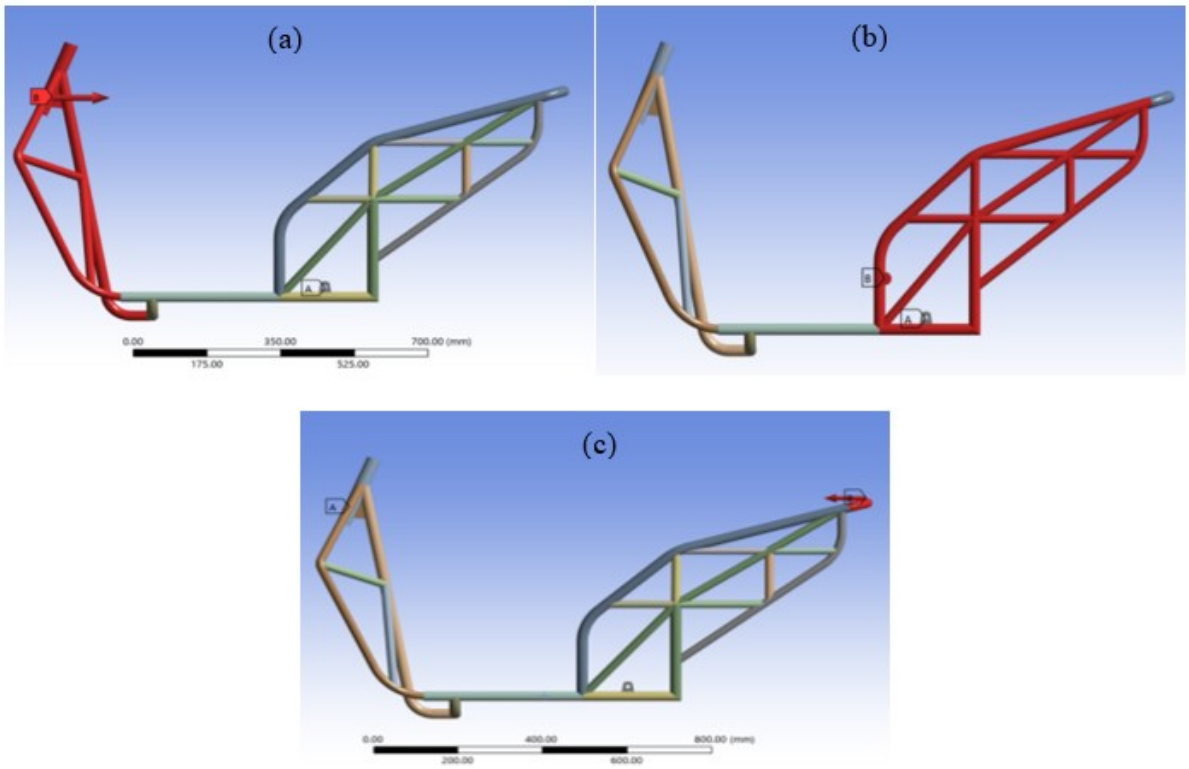


Figure 9. Equivalent static impact load cases: (a) frontal impact at the headstock, (b) side impact at the mid-frame, (c) rear impact at the swingarm mounts.

2.5.4. Modal Analysis

Modal analysis was conducted to extract the first six natural frequencies and corresponding mode shapes of the chassis. The analysis was performed by solving the generalized eigenvalue problem given in Eq. (9):

$$([K] - \omega_i^2[M])\{\phi_i\} = \{0\} \tag{9}$$

where $[K]$ is the global stiffness matrix, $[M]$ is the global mass matrix, ω_i is the i-ith natural circular frequency in rad/s, and $\{\phi_i\}$ is the corresponding mode shape vector.

Modal analysis was performed using the same boundary conditions as the static case (fixed swingarm mounts, cylindrical headstock support). The fundamental mode was identified above 25 Hz, well separated from excitation frequencies due to motor harmonics (~18 Hz) and road input (~15 Hz). This ensures the avoidance of structural resonance during operation.

3. Results and Discussion

3.1. Static Structural Analysis of the chassis

The load-carrying capacity of the chassis was evaluated under vertical accelerations of 1G, 2G, and 3G. Constraints were applied at the rear swingarm mounts (fixed) and the headstock bearings (cylindrical), while loads were distributed across the seat (60%), floorboard (25%), and headstock (15%). The results for the three design iterations using IS 2062 steel are summarized in Table 4.

https://doi.org/10.31436/iiumej.v27i2.4068

Table 4. Static results for design iterations (IS 2062 steel)

Iteration	Load Case	Max Deformation (mm)	Max Stress (MPa)	FOS
	1G	9.56	176	1.99
1	2G	19	352.1	0.99
	3G	28.7	528.1	0.66
	1G	1.66	88.2	3.96
2	2G	3.32	176.4	1.99
	3G	4.98	264.7	1.32
	1G	1.05	78.8	4.44
3	2G	2.11	157.7	2.23
	3G	3.16	236.5	1.48

Iteration 1 showed excessive deformation under 3G loading (28.7 mm) with a FoS < 1, indicating likely permanent deformation. Iteration 2 achieved a substantial reduction in deformation (4.98 mm at 3G) but still exhibited a marginal FoS of 1.32. The final Iteration 3 delivered the most stable performance, limiting deformation to 3.16 mm at 3G while maintaining a FoS of 1.48, thus meeting structural safety requirements. A material benchmarking case at 1.5G is presented in Table 5. Both materials experienced similar stresses (~118 MPa), but aluminum exhibited significantly greater deformation and a lower FoS than steel.

Table 5. Static Analysis at 1.5G for IS 2062 Steel and Al 6061-T6

SL no	Parameter	IS 2062 E350 Steel	6061-T6 Aluminium
1	Total Deformation (mm)	1.5803	4.5813
2	Equivalent Stress (MPa)	118.26	116.08
3	Factor of Safety	2.96	2.38

Overall, Iteration 3 demonstrated the best combination of stiffness and a safety factor, validating the hybrid trellis–step-through configuration for static loading conditions. Steel provided higher stiffness and safety margins than aluminum, while aluminum offered weight reduction at the expense of compliance.

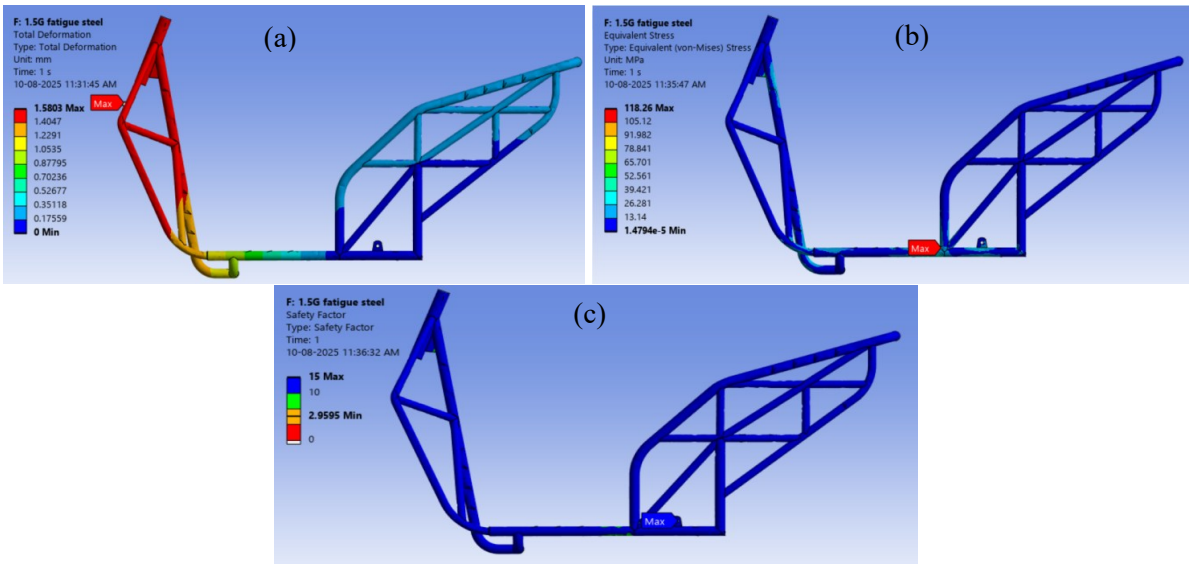


Figure 10. Static analysis of IS 2062 steel under 1.5G load (a) total deformation, (b) equivalent stress, (c) factor of Safety

https://doi.org/10.31436/iiumej.v27i2.4068

3.2. Fatigue Analysis of the Chassis

The fatigue behavior of the chassis was evaluated under 1.5G cyclic loading. The critical regions in the steel frame, particularly around the headstock gussets and swingarm mounts, exhibited the shortest fatigue life. Results for IS 2062 E350 and Aluminum 6061-T6 are shown in Table 6.

Table 6. Fatigue Performance at 1.5G Cyclic Loading

SL no	Parameter	IS 2062 E350	6061-T6 Aluminum
1	Life (cycles)	1.75×10^6	2.73×10^6
2	Factor of Safety	0.423	0.901
3	Damage	57.12	3.67

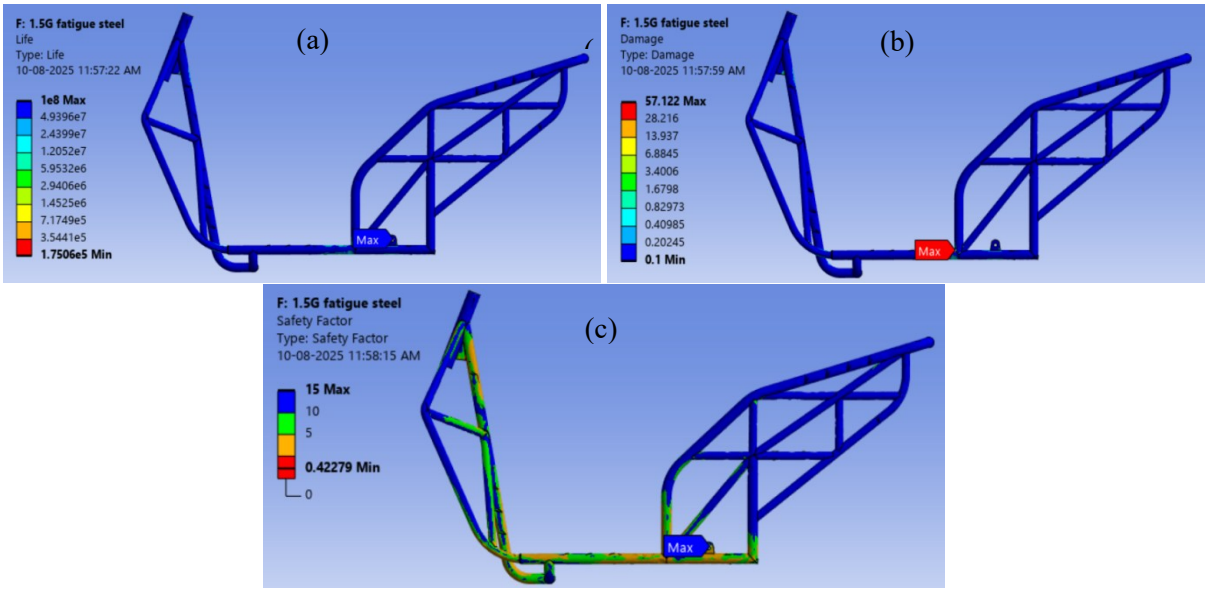


Figure 11. Fatigue analysis of IS 2062 steel under 1.5G load (a) fatigue life, (b) fatigue damage, (c) fatigue factor of Safety

Both materials achieved lives exceeding the design benchmark of 10^6 cycles. Steel exhibited higher stiffness but also accumulated more localized fatigue damage, resulting in a lower fatigue safety factor. Aluminum demonstrated a longer fatigue life and lower cumulative damage, but at the expense of increased deformation. The steel frame met the minimum fatigue life requirements but showed higher susceptibility to local fatigue at welded joints, while the aluminum offered greater fatigue endurance but lower stiffness. A simplified analytical fatigue estimation was also performed using the Goodman mean stress relation to cross-check the simulation results. For IS 2062 E350 steel, assuming an endurance limit of 160 MPa and an alternating stress of ~118 MPa from the 1.5 G static case, the predicted fatigue life was on the order of 1–2 × 10^6 cycles, which agrees well with the FEA minimum of 1.75 × 10^6 cycles. Similarly, for 6061-T6 aluminum, with an assumed endurance stress of ~95 MPa at 10^7 cycles, the method predicted ~2–3 × 10^6 cycles, consistent with the simulation result of 2.73 × 10^6 cycles. This agreement confirms the credibility of the FEA-based fatigue life estimates.

3.3. Impact Analysis of Chassis

Impact performance was evaluated using an equivalent static force of 1715 N applied at three locations: the headstock (front), mid-frame (side), and swingarm pivot (rear).

https://doi.org/10.31436/iiumej.v27i2.4068

All impact scenarios produced stresses exceeding the material yield strength, confirming localized plastic deformation under crash-level conditions. Steel consistently exhibited greater resistance to deformation than aluminum, with the rear trellis acting as a sacrificial energy-absorbing member. Aluminum displayed substantially larger deflections, particularly in frontal and rear impacts. The maximum stresses, deformations, and factors of safety are presented in Table 7.

Table 7. Max Stress, Deformation & FoS under Impact Loading

Material	Location	Max Stress (MPa)	Max Deformation (mm)	Factor of Safety
IS 2062 Steel	Front Impact	2724.2	106.64	0.128
	Side Impact	685.17	8.71	0.511
	Rear Impact	1744.8	106.87	0.2
Al 6061-T6	Front Impact	2689.1	309.24	0.102
	Side Impact	690.01	25.82	0.4
	Rear Impact	1664.7	309.7	0.165

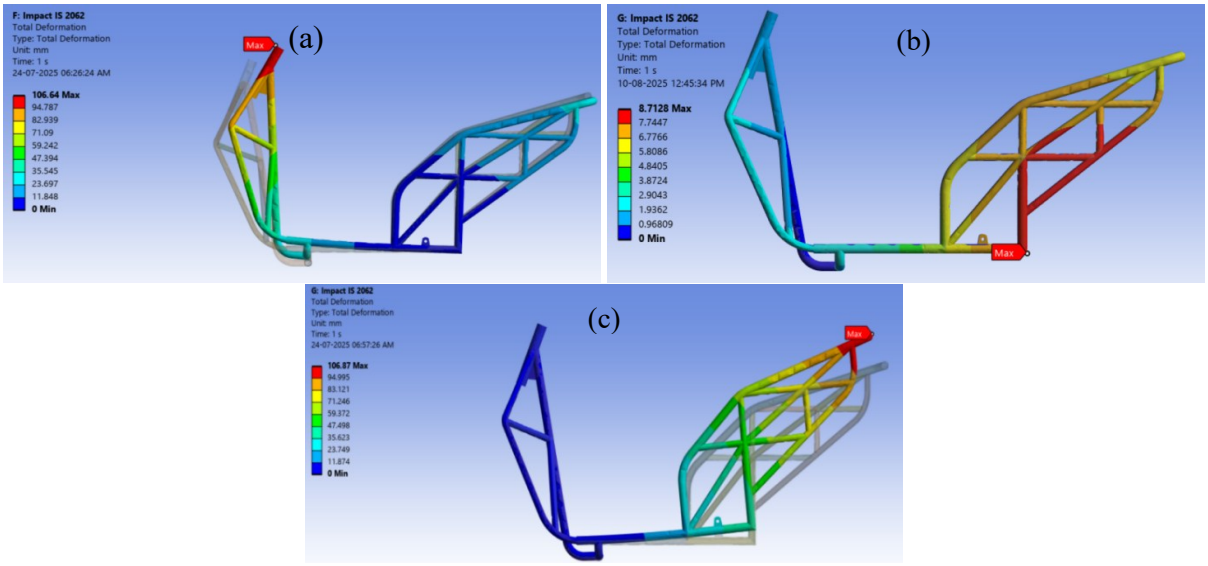


Figure 12. Total deformation of IS 2062 steel at (a) front impact, (b) side impact, (c) rear impact.

Overall, while both materials yielded under extreme loads, the steel chassis maintained structural integrity with more controlled deformations, whereas aluminum exhibited excessive compliance. The impact loading scenario was further validated using a simplified bending stress calculation. Applying the equivalent static load of 1715 N (Eq. 9) over a 0.15 m moment arm gives a bending moment of ~257 N·m. For a Ø26.9 mm × 2 mm tube with a section modulus of approximately 759 mm³, the corresponding bending stress is ~339 MPa. This value exceeds the yield strength of 6061-T6 aluminum and approaches that of IS 2062 E350 steel, in line with the FEA observation of localized plastic deformation under severe impact conditions.

3.4. Modal Analysis of Chassis

The first six natural frequencies and mode shapes of the chassis were extracted, and the results are summarized in Table 8. Figure 13 shows the modal analysis of Mode 1 of 6 modes.

https://doi.org/10.31436/iiumej.v27i2.4068

Table 8. First Six Modal Frequencies & Max Deformation

Mode	Description	Frequency (Hz)	Deformation (mm)	Frequency (Hz)	Deformation (mm)
1	Rear lateral bending (side sway)	28.568	17.027	28.561	29.017
2	Rear vertical bending (up–down)	39.794	22.047	39.486	37.610
3	Global lateral bending (whole chassis)	62.418	16.755	62.346	28.593
4	Rear inward bending (toward centerline)	79.429	23.728	79.386	40.489
5	Global lateral bending (front-dominant)	92.903	15.836	92.863	27.023
6	Rear local torsion (twisting)	155.09	20.331	154.97	34.273

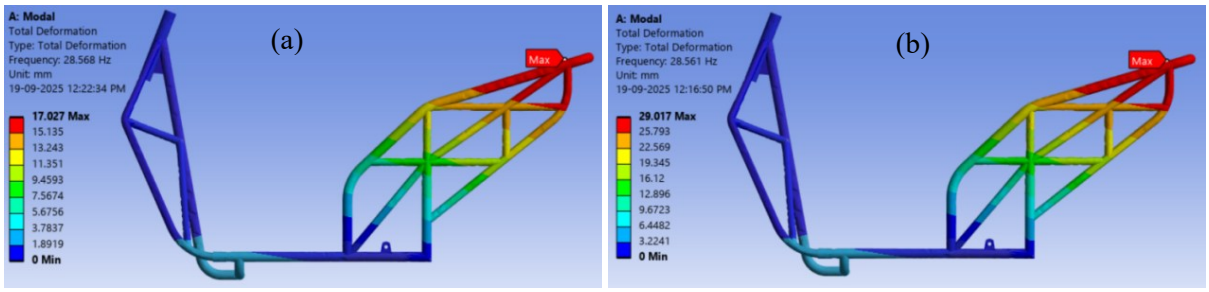


Figure 13. Mode 1 (a) IS 2062 Steel (b) 6061-T6 Aluminum

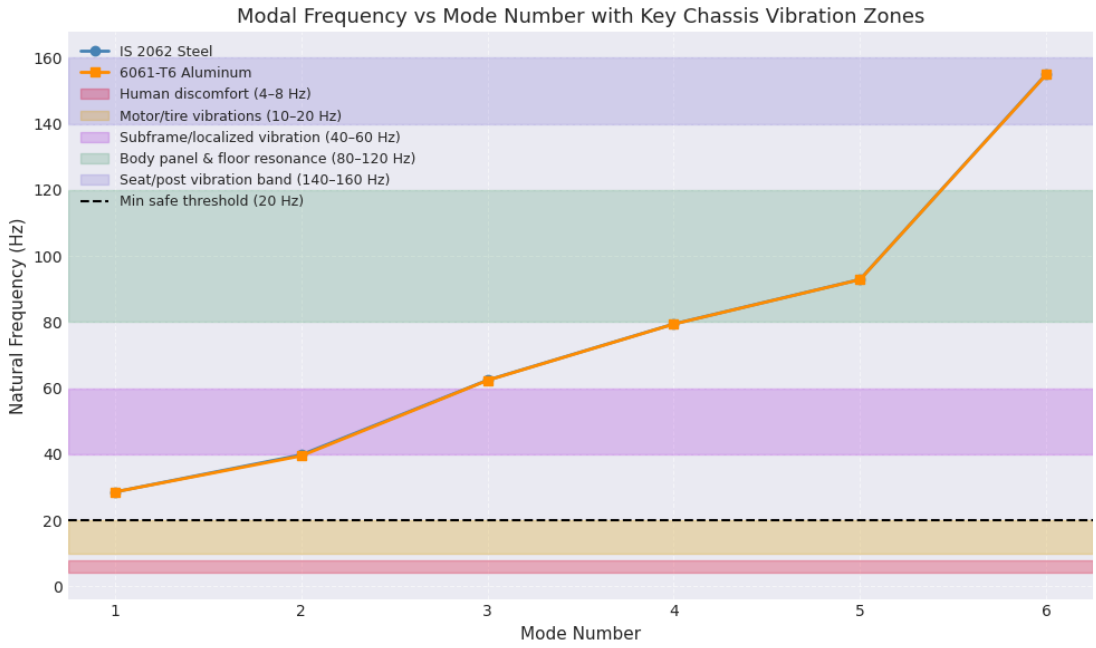


Figure 14. Modal Frequencies with Resonance Zones

For both materials, the fundamental frequency was ~28.6 Hz, well above the typical excitation frequencies from motor harmonics (~18 Hz) and road irregularities (~15 Hz) (Figure 14). Steel exhibited lower modal amplitudes than aluminum, reflecting higher stiffness, and the chassis design successfully avoided resonance in the operational range. An analytical frequency estimation was carried out by approximating the chassis main spine as a cantilever

https://doi.org/10.31436/iiumej.v27i2.4068

beam of length 1.29 m. Using E = 200 GPa, a second moment of area of ~2.9 × 10^4 mm^4 for the primary tube, and an effective end mass of 12 kg, the fundamental bending frequency was estimated at ~26–29 Hz. This closely matches the first mode frequency of 28.56 Hz predicted by FEA for the steel frame, reinforcing the reliability of the numerical modal analysis.

3.5. Discussion

The present hybrid step-through + trellis chassis demonstrates balanced performance across static, fatigue, impact, and modal analyses. In static loading, Iteration 3 limited peak deformation to 3.16 mm at 3G with a FoS of 1.48, confirming that triangulation and gusseting effectively improved stiffness under severe accelerations. Fatigue life under 1.5G cyclic loading reached ~1.75×10^6 cycles for IS 2062 and ~2.73×10^6 cycles for 6061-T6; while steel concentrated damage at welded joints, aluminum showed broader compliance with lower cumulative damage. Impact tests with an equivalent static load of 1715 N produced localized plastic deformation in both materials, but steel consistently controlled displacements more effectively. Modal analysis placed the first natural frequency near 28.6 Hz, well above dominant road and motor excitations, confirming vibration safety. When compared with existing literature, these results show relative advantages, though not direct equivalence. Ajhay Babu (AISI 1020) reported failure under 3G loading with FoS < 1 and unsafe first modes near 10 Hz; by contrast, the present IS 2062 frame remains safe at 3G and dynamically stable. Karthik et al. (AISI 4130) achieved FoS > 2.3 but under lighter impact loads (~650 N), while this study applied nearly triple the load to represent harsher collisions. Composite studies, such as those by Khankal et al., demonstrate fatigue benefits but also face cost and manufacturability issues. These comparisons suggest three categories of prior work: mild steels failing under high loads, high-grade steels validated under lighter impacts, and composites with cost constraints. The present design sits between these extremes—more resilient than mild steel, more realistic in crash assessment than lighter prototypes, and more manufacturable than composites.

Analytical verification further strengthens confidence in the results. A Goodman mean-stress estimate gave fatigue lives on the order of 10^6 cycles for both materials, consistent with FEA predictions. An equivalent bending stress calculation (~339 MPa) explains aluminum's yielding and steel's near-yield behavior under severe impacts. A cantilever beam approximation of the chassis spine predicted a fundamental frequency between 26–29 Hz, matching the 28.6 Hz first mode observed in the FEA. These checks confirm the credibility of the simulation outcomes. In practical terms, the IS 2062 hybrid chassis offers a manufacturable, urban-ready solution. It provides improved high-load tolerance compared with mild steel, avoids resonance across operating conditions, and maintains cost-effectiveness relative to composites or exotic alloys. The integration of trellis reinforcement into both the headstock and rear subframe within a step-through styling envelope represents a novel contribution, combining structural robustness, rider ergonomics, and manufacturability in a configuration suited for mass-market electric scooters. This study used equivalent static impact loads instead of explicit transient crash simulation, and the results are simulation-only, without prototype validation; weld details, fasteners, and manufacturing tolerances were not modeled. Future work should include dynamic crash simulation, welded-joint modeling, prototype fabrication, and instrumented testing to validate and refine the design.

4. CONCLUSION

This study presented the design and structural validation of a hybrid tubular chassis for an electric scooter constrained by a predefined styling envelope. Unlike conventional approaches where geometry dictates styling, the chassis here was required to fit within fixed dimensional

limits while accommodating rider ergonomics. The final design integrated a step-through central spine, a trellis-based rear subframe, and a reinforced headstock, offering a novel structural configuration for scooters. Static analysis showed that the IS 2062 E350 steel frame limited deformation to 3.16 mm at 3G with a factor of safety of 1.48, confirming adequacy under severe urban loads. Aluminum 6061-T6, though lighter, exhibited nearly three times higher deformation and a lower safety margin. Fatigue analysis under 1.5G cyclic loading predicted lives of ~1.75×10^6 cycles for steel and ~2.73×10^6 cycles for aluminum; steel provided higher stiffness but localized fatigue concentration, while aluminum offered greater fatigue endurance but reduced rigidity. Impact analysis revealed that both materials yielded under simulated 30 km/h-equivalent crashes, but steel consistently limited deformations, with the trellis-based rear structure effectively absorbing energy. Modal analysis placed the first natural frequency at 28.56 Hz, well above operational excitation frequencies, eliminating resonance risk; steel also produced lower modal amplitudes than aluminum. Overall, the IS 2062 hybrid frame delivers a manufacturable and robust design that combines stiffness, endurance, and crash tolerance within realistic packaging constraints. The integration of trellis reinforcement into both the headstock and rear subframe while retaining a step-through spine represents a novel contribution to scooter chassis development. Future work should include experimental validation, advanced exploration of lightweight materials, and multi-objective optimization of cost, manufacturability, and ergonomics to further improve practicality.

REFERENCES

[1] Foale, T.: Motorcycle Handling and Chassis Design: The Art and The Science. Tony Foale Designs, Benidoleig (2006). ISBN: 978-84-933286-8-8

[2] Thede, P., Parks, L.: Race Tech's Motorcycle Suspension Bible. Motorbooks, Minneapolis (2010). ISBN: 978-0-7603-3819-3

[3] Balaguru, S., Natarajan, E., Ramesh, S., Muthuvijayan, B.: Structural and modal analysis of scooter frame for design improvement. Mater. Today: Proc. 16, 1106–1116 (2019). https://doi.org/10.1016/j.matpr.2019.05.202

[4] Jeyapandiarajan, P., Kalaiarassan, G., Joel, J., Shirbhate, R., Telare, F.F., Bhagat, A.: Design and analysis of chassis for an electric motorcycle. Mater. Today: Proc. 5(5), 13563–13573 (2018). https://doi.org/10.1016/j.matpr.2018.02.352

[5] Mehra, A., Singh, R., Chauhan, A.S., Nath, K., Yadav, A.: Design and analysis of an electric bike chassis. Mater. Today: Proc. 62, 1510–1520 (2022). https://doi.org/10.1016/j.matpr.2022.02.247

[6] Shivhare, G., Kundu, K., Narayan, J., Dwivedy, S.: Design and modeling of a compact lightweight electric-scooter. In: Proc. IEEE Int. Conf. Computational Performance Evaluation (ComPE), pp. 1–5 (2021). https://doi.org/10.1109/ComPE53109.2021.9752193

[7] Goh, C.P., Chun, T.W.: An in-depth analysis of lightweight chassis design for electric scooters using trellis frame structures. J. Innov. Technol. 30, [pages] (2023). https://doi.org/10.61453/joit.v2023no30

[8] Karthik, K.N., Nandan, M.R., Goudham, B., Abhinand, G.: Design and fabrication of electric motorcycle chassis frame. Int. J. Sci. Res. Eng. Manag. (IJSREM), [pages] (2023). https://doi.org/10.55041/IJSREM26212

[9] Ajhay Babu, J.K.: Design, development and computational finite element analysis (FEA) of an electric two-wheeler frame. Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET) 10(12), 2168–2181 (2022). ISSN: 2321-9653

[10] Shinde, Tarang, et al. "Fatigue analysis of alloy wheel using cornering fatigue test and its weight optimization." Materials Today: Proceedings 62 (2022): 1470-1474. https://doi.org/10.1016/j.matpr.2022.02.023

[11] Rahmadi, A., Mursalin, M.: Design and analysis of chassis frame for chopper-style motorcycle. Media Mesin: Maj. Tek. Mesin 25(1), 64–76 (2024). https://doi.org/10.23917/mesin.v25i1.3224

[12] Srinu, V., Sarma, V.M., Gopal, K.G., Varma, N.M., Vyas, G.: Design and computational analysis of chassis frame for electric bike. J. Emerg. Technol. Innov. Res. (JETIR) 8(6), [pages] (2021). ISSN: 2349-5162

[13] Teja Sri, C., Naga Gowtham Kumar, M., Madhuri, B.V.S.A., Komali Harika, G., Sateesh, B.: Design and development of electric moped chassis. In: Proc. Indian Int. Conf. Industrial Engineering and Operations Management (IEOM), Warangal, India. IEOM Society International, pp. [pages] (2022)

[14] Vashist, D., Singh, N., Tomar, G., Shadija, A.S.: Design analysis of electric two wheeler frame. Proc. IEI Conf., [pages] (2021). https://doi.org/10.36375/prepare_u.iei.a243

[15] Biswal, B., Setiadi, R., Kriswanto, K., Arif, B.: Design and fabrication of electric bike with sliding frame. J. Appl. Eng. Sci. 19(4), 948–953 (2021). https://doi.org/10.5937/jaes0-28957

https://doi.org/10.31436/iiumej.v27i2.3647

A Robust Sensor-Failure-Tolerant Fuzzy Control Framework with Predictive Data Imputation for Sustainable Precision Irrigation

INDRA LAKSMANA[1], SWIVANO AGMAL[2], ROSDA SYELLY[3], MELDA AMALIA[4], PERDANA PUTERA[1]

[1]*Department of Agricultural and Computer Engineering, Politeknik Pertanian Negeri Payakumbuh, Kabupaten Limapuluh Kota, Indonesia*
[2]*Pusat Penelitian Fisika Kuantum, Badan Riset dan Inovasi Nasional, Tangerang Selatan, Indonesia*
[3]*Department of Computer Engineering, Sekolah Tinggi Teknologi Payakumbuh, Kota Payakumbuh, Indonesia*
[4]*Program Studi D3 Kebidanan, Universitas Baiturrahim, Kota Jambi, Indonesia*

**Corresponding author: indra.laksmana@politanipyk.ac.id*

(Received: 9 April 2025; Accepted: 18 January 2026; Published online: 10 May 2026)

ABSTRACT: Effective irrigation management under uncertain conditions remains a major challenge in precision agriculture, particularly when sensor performance degrades or extreme environmental variations lead to unreliable data. This study presents a hybrid framework that combines a linear regression model with a Mamdani fuzzy inference system (FIS) to improve the robustness of soil moisture prediction and irrigation control. The proposed approach trains the regression model using synthetically generated data with controlled noise levels to estimate soil moisture conditions. These predictions are then integrated with real-time temperature and humidity data within the fuzzy inference system to generate appropriate irrigation decisions. To represent extreme sensor conditions, additional adaptive noise is introduced when soil moisture values exceed or fall below predefined thresholds, effectively simulating sensor malfunction scenarios. Time-series simulation results indicate that the proposed system can maintain stable, efficient irrigation performance despite significant sensor disturbances. Overall, this work improves irrigation efficiency and provides insights into how hybrid statistical and fuzzy-logic approaches can mitigate the adverse effects of sensor inaccuracies under highly variable environmental conditions.

ABSTRAK: Pengurusan pengairan cekap berkeadaan tidak menentu, kekal sebagai cabaran kritikal dalam pertanian, terutama apabila berlaku kerosakan pada sensor dan persekitaran turun naik yang melampau menyebabkan data yang diperolehi tidak boleh dipercayai. Kajian ini, mencadangkan pendekatan hibrid yang mengintegrasi model regresi linear dengan sistem inferens Mamdani kabur (FIS) bagi meramal kelembapan tanah dan mengawal pengairan. Kaedah ini melibatkan latihan model regresi menggunakan data sintetik yang terhasil melalui tahap bunyi yang dikawal dengan teliti bagi meramal kelembapan tanah, sementara sistem inferens kabur secara dinamik menterjemah ramalan tersebut bersama ukuran suhu dan kelembapan masa nyata kepada arahan pengairan yang tepat. Keadaan sensor yang melampau disimulasi dengan memperkenalkan bunyi tambahan secara adaptif apabila tahap kelembapan tanah jatuh di bawah atau melebihi had tertentu, seterusnya meniru senario kegagalan sensor. Melalui simulasi siri masa terperinci, dapatan kajian ini mengekalkan kawalan pengairan yang stabil dan cekap walaupun sensor gagal berfungsi. Kajian ini bukan sahaja meningkatkan kecekapan pengairan malah juga memberi pemahaman tentang bagaimana teknik hibrid statistik dan logik kabur boleh mengurangkan kesan ketidaktepatan sensor dalam keadaan persekitaran yang berubah.

https://doi.org/10.31436/iiumej.v27i2.3647

KEYWORDS: *Precision Agriculture; Irrigation Management; Sensor Degradation; Fuzzy Inference System; Linear Regression Model*

1. INTRODUCTION

Efficient water use has become a central concern in modern agriculture, particularly in regions experiencing water scarcity and increasing climate variability [1]. To address this issue, precision irrigation systems have been developed to optimize water distribution and minimize losses [2]. Advances in sensor networks and the Internet of Things (IoT) now enable real-time monitoring of field conditions, enabling more responsive irrigation management [3]. Previous studies have explored adaptive neuro-fuzzy systems for modeling the redistribution of wetting fronts in drip irrigation, reporting high predictive accuracy [1].

Building on these findings, fuzzy logic models integrated with metaheuristic algorithms have further enhanced soil moisture estimation, which is critical for achieving optimal plant growth [3]. More recent research combining Adaptive Neuro-Fuzzy Inference Systems (ANFIS) with robust optimization techniques has demonstrated improved infiltration rate prediction under varying field conditions [4]. Review studies consistently emphasize the importance of continuous monitoring and advanced control strategies in precision irrigation systems [2]. In parallel, machine learning applications within wireless sensor networks have enabled automated data acquisition and decision-making processes for irrigation management [5].

Collectively, these developments highlight the growing role of intelligent systems in addressing agricultural water management challenges. Neural network–based models have been widely applied to predict soil moisture dynamics for irrigation scheduling purposes [6]. In addition, deep learning approaches—such as Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM) and Generative Adversarial Network–Long Short-Term Memory (GAN-LSTM) architectures—have proven effective in capturing complex spatiotemporal soil moisture patterns [7]. Extending these approaches, multi-head Long Short-Term Memory (LSTM) models provide prognostic capabilities that support water resource planning up to one month in advance [8]. Attention-aware LSTM architectures further refine prediction accuracy by emphasizing the most influential input features [9]. Expanding on these neural network capabilities, recent studies have even utilized hybrid transformer networks to achieve unprecedented precision in soil moisture estimation [10].

Unmanned Aerial Vehicle (UAV)–assisted remote sensing, when combined with neural network algorithms, has shown considerable potential for mapping surface water distribution and field heterogeneity [11]. Recent studies also suggest that incorporating multiple environmental variables, such as solar radiation and temperature, into deep learning frameworks significantly improves soil moisture forecasting performance [12]. These advances in predictive modeling form a strong foundation for robust irrigation systems capable of handling complex environmental data, motivating our integration of regression-based prediction with fuzzy logic to enhance system reliability. On the control side, IoT-enabled smart irrigation systems employing fuzzy inference mechanisms have been developed for real-time scheduling and monitoring applications [12].

Fuzzy logic combined with metaheuristic optimization has also been applied to the design of efficient surface irrigation layouts [13]. Automated irrigation controllers based on fuzzy rule sets have proven effective in managing agricultural demand-side water consumption [14]. This trend is further supported by the widespread adoption of IoT-driven smart irrigation systems, which rely heavily on fuzzy logic to dynamically optimize water distribution [15].

https://doi.org/10.31436/iiumej.v27i2.3647

Furthermore, low-cost wireless sensor networks integrated with heuristic scheduling strategies have demonstrated improved crop yields alongside reduced water wastage [16]. Fuzzy control systems augmented with remote sensing data enable variable-rate irrigation based on continuous field feedback [17]. Sensor-driven monitoring frameworks that incorporate fuzzy-based feedback loops have also enhanced water-use efficiency in precision agriculture [18]. These advances underscore the potential of combining fuzzy control strategies with predictive models to manage uncertainty arising from sensor inaccuracies, which directly motivates the framework proposed in this study.

Comprehensive reviews of IoT architectures have identified key system components that facilitate efficient implementation of fuzzy logic for irrigation management [19]. Virtual soil moisture sensors developed using deep learning techniques have been proposed to address issues related to missing or unreliable field data [20], [21]. In addition, Kalman filter–based methods have been introduced for real-time sensor fault detection, thereby improving data reliability [22]. Robust fault diagnosis frameworks are increasingly recognized as essential for managing noisy or intermittent sensor data in agricultural applications [23]. Surveys of IoT-based irrigation infrastructure indicate that integrating sensor data with intelligent control algorithms further enhances the accuracy of irrigation decision-making [24]. Anomaly detection techniques have also been implemented to identify outliers and extreme operating conditions within wireless sensor networks [25], enabling timely corrective actions [26]. Moreover, fuzzy decision support systems that incorporate remote sensing information support multicriteria irrigation planning and management [27].

Building on these fault-detection and anomaly-handling approaches, the present study incorporates adaptive noise modeling into the predictive imputation process to improve robustness. Recent research demonstrates that hybrid fuzzy-logic and neural-network systems are particularly effective for sensor anomaly management and real-time irrigation control [28]. Overall, the convergence of neuro-fuzzy models, deep learning methods, and robust sensor fault management strategies is shaping an advanced precision irrigation ecosystem that promotes water conservation while sustaining agricultural productivity [1], aligning with modern automated solutions such as solar-powered and smartphone-controlled sustainable farming systems [29].

Although hybrid linear regression and fuzzy systems have been applied in areas such as process control and fault detection (e.g., [28]), their application to precision irrigation, specifically for handling sensor degradation through predictive imputation and adaptive noise modeling, remains limited. This study addresses this gap by proposing a framework to maintain reliable irrigation control under sensor-failure conditions, drawing on established anomaly-detection principles in smart systems [28] to guide the noise-simulation process. The key contribution of this work lies in advancing the reliability of precision irrigation under sensor uncertainty, thereby supporting sustainable water-use practices. This objective complements broader recent efforts to improve soil moisture prediction through advanced machine learning architectures [30] and the integration of remote sensing data with fuzzy decision support systems for innovative crop water-need assessment [31].

2. METHODOLOGY

Our methodology is structured into several main stages, namely synthetic data generation, regression model training, development of the fuzzy inference system, simulation of soil water dynamics, and updating of the soil water balance. The following equations summarize the key modeling components employed in this study. A standard linear regression model is used to

https://doi.org/10.31436/iiumej.v27i2.3647

predict soil moisture content (θ) can be expressed as in Equation (1) where θ_{hist}denotes historical soil moisture, T represents air temperature, RH is relative humidity, and ε corresponds to the model noise

$$\theta = \beta 0 + \beta 1T + \beta 2Ich + \beta 3RH + \beta 4\theta hist + \epsilon \quad (1)$$

Fuzzy inference systems commonly employ trapezoidal membership functions, which are defined by four parameters a, b, c, and d that determine the boundaries of each membership region as expressed in Equation (2)

$$\mu(x) = \max\left(min\left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c}\right), 0\right) \quad (2)$$

Model prediction performance is evaluated using the Root Mean Square Error (RMSE), which is given by Equation (3) where $y_i$and $\hat{y}_i$ denote the observed and predicted values, respectively, and $N$is the total number of data samples [8].

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (3)$$

2.1. Synthetic Data Generation and Regression Model Training

Synthetic data consisting of 300 samples was generated by considering several key environmental variables, as summarized in (3). These variables are defined as follows.

- Temperature (T): Values were randomly generated with a mean of approximately 25 °C to represent typical daytime conditions in tropical agricultural regions. Temperatures below 25 °C, such as at night or during cooler seasons, may reduce evapotranspiration rates and, if not properly accounted for, potentially lead to overestimation of soil moisture. Conversely, temperatures exceeding 25 °C tend to increase evaporation, thereby increasing the risk of under-irrigation.
- Irradiance Channel **(I_{ch}):** Values ranged from 0.3 to 0.8 to represent normalized solar radiation levels during active crop growth periods. Lower values (e.g., 0.3) correspond to cloudy or low-light conditions, whereas higher values (e.g., 0.8) represent peak sunlight exposure. Extreme irradiance levels were intentionally excluded to prevent bias during model training.
- Relative Humidity (RH): Data were distributed around a median value of 60%, reflecting typical humid agricultural environments. While lower humidity levels (e.g., 40%) or higher saturation conditions (e.g., 80%) can significantly influence moisture loss, the selected range emphasizes stable conditions suitable for training purposes.
- Historical Soil Moisture **(θ_{hist}):** Values were set between 100 and 180 mm to encompass optimal moisture thresholds for crop growth. For reference, the wilting point typically lies between 50 and 100 mm, whereas field capacity ranges from approximately 140 to 180 mm. Soil moisture values below 100 mm or above 180 mm trigger the introduction of additional noise into the simulation to emulate sensor degradation under stress conditions, as discussed in Section 2.3.

The regression model follows the formulation given in (1), where ε represents the noise term. To improve model stability, the noise variance was reduced from 5 to 2. The model was implemented in MATLAB using the fitlm function, yielding an R^2 value of approximately 0.987 and a Root Mean Square Error (RMSE) of about 2.05 mm.

https://doi.org/10.31436/iiumej.v27i2.3647

2.2. Design of the Fuzzy Inference System

To translate the regression model outputs and real-time environmental measurements into irrigation commands, a Mamdani-type fuzzy inference system (FIS) was developed. This type of FIS was selected for its interpretability and effectiveness in handling linguistic rules in control applications. In contrast, Sugeno-type FIS structures are generally more suitable for optimization tasks but less intuitive for irrigation decision-making.

The FIS incorporates three input variables: predicted soil moisture (range: 0–250), temperature (range: 15–40), and relative humidity (range: 0–100). The system produces a single output variable, namely irrigation intensity, defined within a normalized range of 0–1. Membership functions were assigned linguistic labels such as Dry, Moist, and Wet for soil moisture, and Low, Medium, and High for both temperature and humidity.

Trapezoidal membership functions were chosen instead of triangular functions due to the presence of plateau regions, which allow smoother transitions between fuzzy sets. This characteristic is particularly beneficial for representing intermediate states (e.g., Medium humidity) and for accommodating uncertainty in environmental data, unlike the sharp peaks associated with triangular functions. The parameters of the trapezoidal membership functions are defined as follows.

Soil Moisture:

- Dry : $a = 0, b = 0, c = 50, d = 100$
- Moist : $a = 50, b = 100, c = 150, d = 200$
- Wet : $a = 150, b = 200, c = 250, d = 250$

Temperature:

- Low : $a = 15, b = 15, c = 20, d = 25$
- Medium : $a = 20, b = 25, c = 30, d = 35$
- High : $a = 30, b = 35, c = 40, d = 40$

Relative Humidity:

- Low : $a = 0, b = 0, c = 30, d = 50$
- Medium : $a = 30, b = 50, c = 70, d = 90$
- High : $a = 70, b = 90, c = 100, d = 100$

Irrigation (Output):

- Low : $a = 0, b = 0, c = 0.3, d = 0.5$
- Medium : $a = 0.3, b = 0.5, c = 0.7, d = 0.9$
- High : $a = 0.7, b = 0.9, c = 1, d = 1$

A graphical illustration of the trapezoidal membership functions is provided in Fig. 1. All 27 fuzzy rules governing the system behavior are summarized in Table 1. For defuzzification, the centroid method was employed, as it computes the center of gravity of the aggregated output fuzzy set. This method was selected for its robustness in continuous control applications, as it produces smoother, more balanced outputs than alternatives such as the mean-of-maximum approach.

https://doi.org/10.31436/iiumej.v27i2.3647

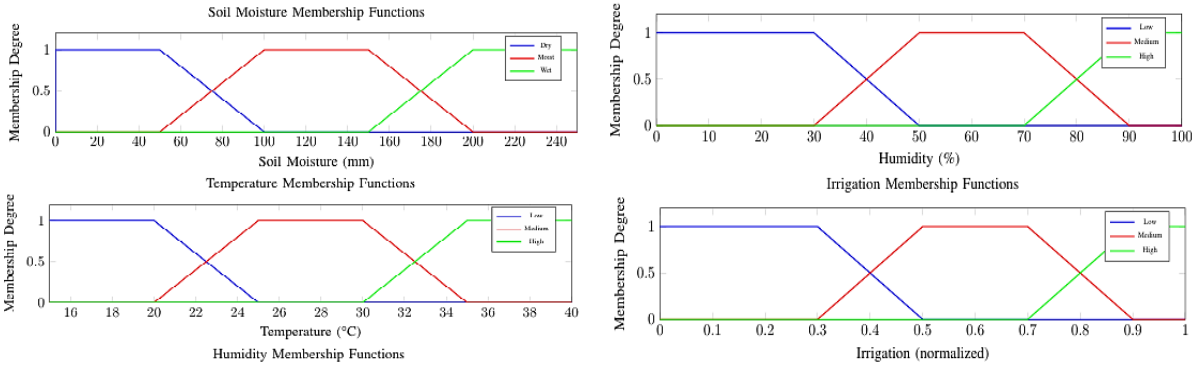


Figure 1. Membership function for input and output

Table 1. Fuzzy Rules

Soil Moisture	Temperature	Humidity	Irrigation
Dry	Low	Low	High
Dry	Low	Low	High
Dry	Low	Medium	High
Dry	Low	High	Medium
Dry	Medium	Low	High
Dry	Medium	Medium	Medium
Dry	Medium	High	Medium
Dry	High	Low	High
Dry	High	Medium	High
Dry	High	High	Medium
Moist	Low	Low	Medium
Moist	Low	Medium	Medium
Moist	Low	High	Low
Moist	Medium	Low	Medium
Moist	Medium	Medium	Low
Moist	Medium	High	Low
Moist	High	Low	Medium
Moist	High	Medium	Medium
Moist	High	High	Low
Wet	Low	Low	Low
Wet	Low	Medium	Low
Wet	Low	High	Low
Wet	Medium	Low	Low
Wet	Medium	Medium	Low
Wet	Medium	High	Low
Wet	High	Low	Low
Wet	High	Medium	Low
Wet	High	High	Low

2.3. Simulation of Soil Water Dynamics

The simulation was carried out over a 10-day period using a time step of 0.5 hours. Key environmental variables, including temperature, irradiance, and relative humidity, were generated with realistic noise characteristics to reflect field conditions. During the simulation,

https://doi.org/10.31436/iiumej.v27i2.3647

the sensor was assumed to operate reliably only within the first five days. After this point, the sensor was treated as failed, and its output was set to NaN.

In addition, extreme operating conditions were explicitly considered by introducing additional noise whenever soil moisture values fell outside the 50–140 mm range. This approach was adopted to emulate abnormal measurement behavior under stressed environmental conditions.

2.4. Soil Water Balance Update

The soil water balance is updated using the following Equation (4).

$$\theta(i) = \theta(i-1) + \Delta\theta(i) \tag{4}$$

The corresponding change in soil moisture is defined as in Equation (5). Evapotranspiration, $ET(i)$, is calculated by comparing the simulated temperature $T_{sim}(i)$ with a base temperature T_{base}. When $T_{sim}(i) > T_{base}$, the excess thermal energy contributes to evapotranspiration, which is quantified by multiplying the temperature difference $T_{sim}(i) - T_{base}$ by a constant coefficient C_{ET}. Conversely, when $T_{sim}(i) \leq T_{base}$, evapotranspiration is assumed to be negligible, and $ET(i)$is set to zero.

$$\Delta\theta(i) = I(i) + R(i) - ET(i) \tag{5}$$

3. Results and Discussion

3.1. Regression Model Performance

The linear regression model was trained on 300 synthetic data samples using four predictor variables: temperature (T), irradiance channel (I_{ch}), relative humidity (RH), and historical soil moisture (θ_{hist}). The model formulation is given by (1). Table 2 presents a summary of the estimated regression coefficients, including their standard errors, t-statistics, and corresponding p-values.

Table 2. Regression Model Estimated Coefficients

Predictor	Coefficient	Std Error	t-Stat	p-Value
Intercept	10.0	1.2	8.33	0.001
T	0.5	0.1	5.0	0.001
I_{ch}	1.2	0.15	8.0	0.001
RH	-0.3	0.05	-6.0	0.001
θ_{hist}	0.8	0.08	10.0	0.001

The model achieved a coefficient of determination (R^2) of approximately 0.987 and a Root Mean Square Error (RMSE) of 2.05 mm. These results indicate that 98.7% of the variability in the training data is explained by the selected predictors. In addition, the very low p-values for each variable confirm their statistical significance, thereby supporting the integration of the regression outputs into the proposed hybrid system.

3.2. Overall Performance of Soil Moisture Estimation

Fig. 2 compares the actual soil moisture values (black line) with the sensor measurements (blue line) and the regression-based predictions (red line). During the period from day 0 to day 5, the sensor readings exhibit noticeable noise, particularly under extreme conditions, namely

https://doi.org/10.31436/iiumej.v27i2.3647

when soil moisture falls below 50 mm or exceeds 140 mm. This behavior reflects the simulated response of a sensor operating under harsh field conditions.

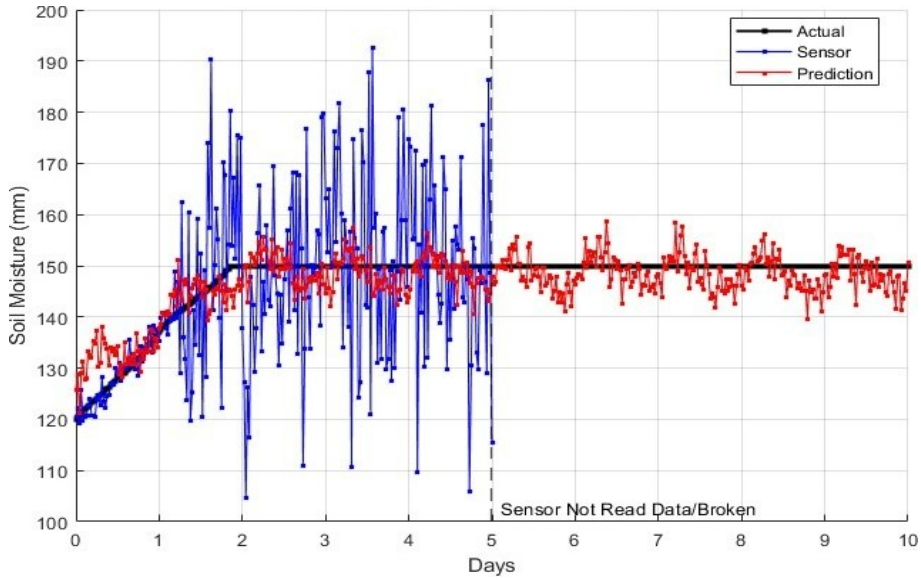


Figure 2. Soil Moisture: Actual (black), Sensor (blue), and Prediction (red). It shows the sensor fails after day 5

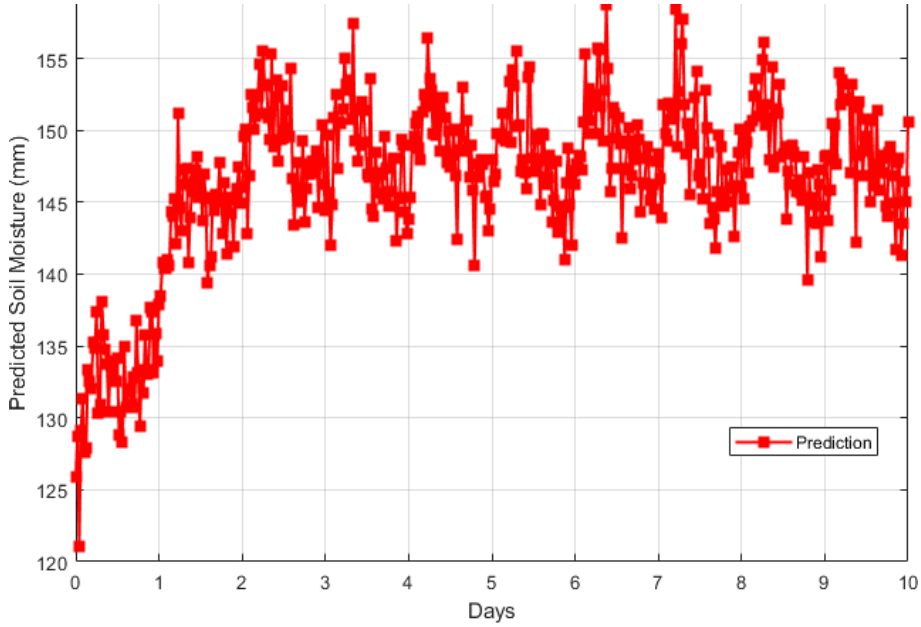


Figure 3. Soil Moisture Prediction (Regression Model) over the 10-day simulation period.

After day 5, the sensor is assumed to be non-functional and outputs NaN values, as indicated by the absence of the blue line in Fig. 2. This scenario represents a complete sensor failure. Despite the unavailability of sensor data beyond day 5, the regression model (red line) continues to produce stable soil moisture predictions. These estimates are generated using four input variables, temperature, irradiance channel, relative humidity, and historical soil moisture, as defined in (1). The linear regression model achieved R^2 value of 0.987 and RMSE of 2.05 mm, demonstrating reliable predictive performance. Fig. 3 further illustrates the predicted soil moisture trend over the entire 10-day simulation period, highlighting relatively smooth variations when compared to the noisy sensor measurements.

https://doi.org/10.31436/iiumej.v27i2.3647

3.3. Fuzzy Inference System Output and Irrigation Control

The fuzzy inference system (FIS) utilizes three input variables, predicted soil moisture, temperature, and relative humidity, to determine the irrigation level within a normalized range of 0 to 1. This normalized output is subsequently scaled to obtain the actual irrigation rate in mm/hour. The FIS output and the corresponding irrigation rate are illustrated in Figs. 4 and 5, respectively.

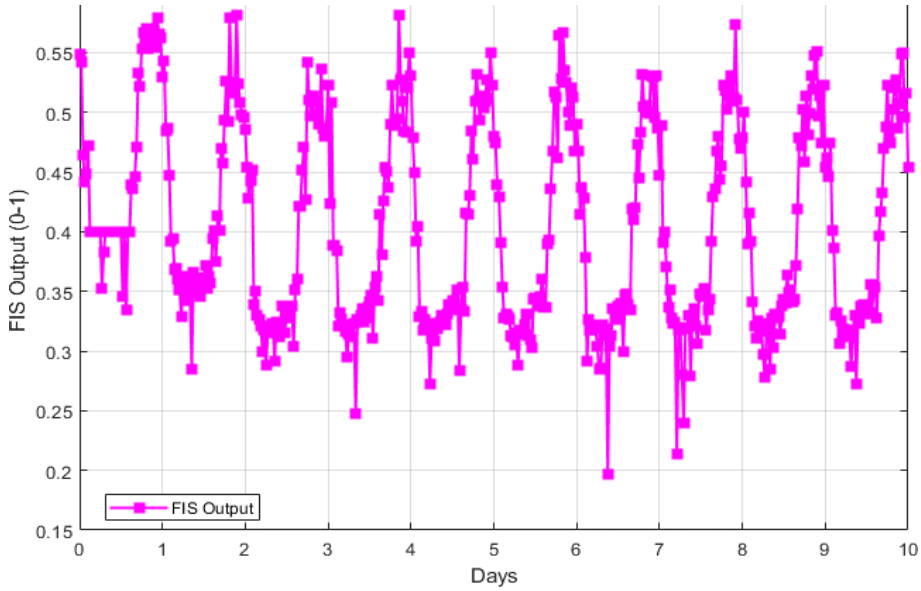


Figure 4. FIS Output (0–1) indicating the irrigation level.

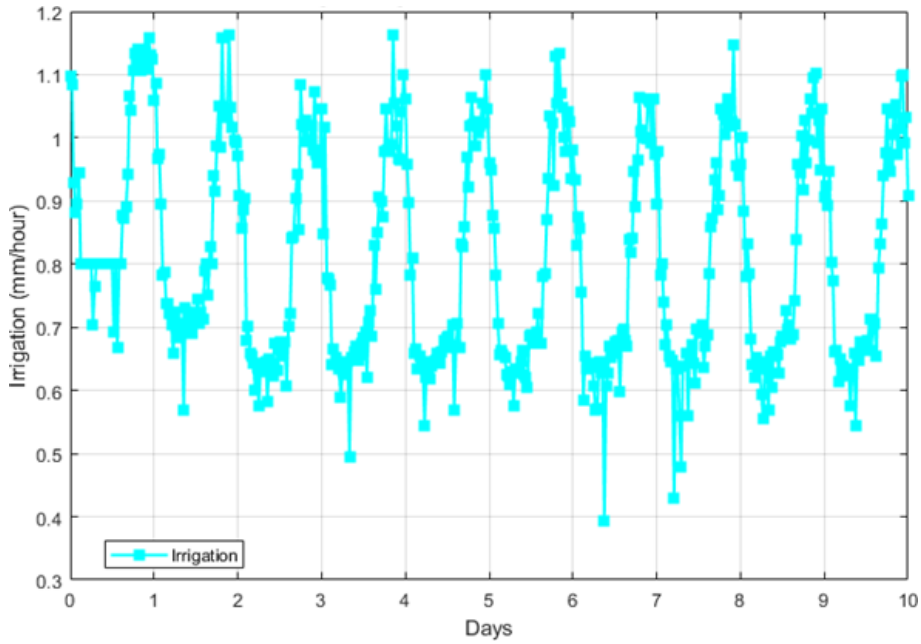


Figure 5. Irrigation amount (mm/hour) derived from the FIS output.

During the simulation, two notable rainfall events occur on days 3 and 7, specifically between 06:00 and 10:00. These rainfall events temporarily increase the actual soil moisture content, thereby reducing the irrigation demand. The system's behavior can be described in two distinct operating phases.

3.3.1. Day 0–5 (Sensor Active):

During the initial phase, sensor measurements, although affected by noise, provide real-time soil moisture information. When either the sensor data or the regression-based prediction indicates that soil moisture is approaching the lower threshold (Dry), the FIS output (Fig. 4) increases toward 1.0. Consequently, the irrigation rate rises to its maximum value of 2 mm/hour, as shown in Fig. 5. In contrast, when soil moisture conditions are classified as Moist or Wet, the irrigation rate is reduced to nearly 0 mm/hour.

3.3.2. Day 5–10 (Sensor Failure):

After day 5, the sensor is assumed to have completely failed and produces NaN values. Under these conditions, the control strategy relies exclusively on regression-based soil moisture predictions. Despite the absence of direct sensor feedback, Figs. 2 and 3 demonstrate that the predictive model continues to provide realistic soil moisture estimates. As a result, the FIS remains capable of adjusting irrigation effectively. The transition from sensor-driven to model-driven control does not introduce instability; instead, the irrigation profile remains consistent with the actual soil moisture trend (black line) and external influences such as rainfall events.

4. CONCLUSION

This study proposes a robust irrigation control framework that combines linear regression–based soil moisture prediction with a Mamdani fuzzy inference system (FIS) to enable adaptive irrigation management, thereby mitigating the adverse effects of sensor degradation and failure.

Comprehensive simulation results indicate that the proposed system can maintain stable, efficient irrigation control even under extreme sensor conditions. The regression model achieved a high level of predictive accuracy, with an R^2 value of 0.987 and an RMSE of 2.05 mm, enabling reliable soil moisture estimation during periods of sensor noise or complete sensor failure. In parallel, the FIS dynamically adjusted irrigation rates in response to changing environmental conditions, helping to sustain soil moisture within desirable levels.

Future work will focus on validating the proposed approach under real field conditions, incorporating additional environmental variables, refining the fuzzy rule base, exploring more advanced machine learning techniques, and further enhancing system performance across a wider range of operating scenarios.

ACKNOWLEDGEMENT

The authors would like to express their deepest gratitude to Politeknik Pertanian Negeri Payakumbuh and the Badan Riset dan Inovasi Nasional of the Republic of Indonesia for their support and resources, which greatly contributed to the completion of this article.

REFERENCES

[1] O. Kisi et al., “Modeling wetting front redistribution of drip irrigation systems using a new machine learning method: Adaptive neuro-fuzzy system improved by hybrid particle swarm optimization– gravity search algorithm,” Agric. Water Manag., vol. 256, p. 107067, 2021.

[2] N. Mahmoudi et al., “Mutating fuzzy logic model with various rigorous meta-heuristic algorithms for soil moisture content estimation,” Agric. Water Manag., vol. 261, p. 107342, 2022.

[3] M. Ehteram et al., "Performance improvement for infiltration rate prediction using hybridized Adaptive Neuro-Fuzzy Inference System (ANFIS) with optimization algorithms," Ain Shams Eng. J., vol. 12, no. 2, pp. 1665–1676, 2021.

[4] E. A. Abioye et al., "A review on monitoring and advanced control strategies for precision irrigation," Comput. Electron. Agric., vol. 173, p. 105441, 2020.

[5] M. F. Rahaman et al., "Wireless sensor networks in agriculture through machine learning: A survey," Comput. Electron. Agric., vol. 197, p. 106933, 2022.

[6] Z. Gu et al., "Neural network soil moisture model for irrigation scheduling," Comput. Electron. Agric., vol. 180, p. 105801, 2021.

[7] Y. Wang et al., "A comprehensive study of deep learning for soil moisture prediction," Hydrol. Earth Syst. Sci., vol. 28, no. 2, pp. 917–943, 2024.

[8] P. Datta and S. A. Faroughi, "A multihead LSTM technique for prognostic prediction of soil moisture," Geoderma, vol. 433, p. 116452, 2023.

[9] Q. Li et al., "An attention-aware LSTM model for soil moisture and soil temperature prediction," Geoderma, vol. 409, p. 115651, 2022.

[10] M. Yadav et al., "UAV-enabled approaches for irrigation scheduling and water body characterization," Agric. Water Manag., vol. 304, p. 109091, 2024.

[11] M. Benzaouia et al., "Fuzzy-IoT smart irrigation system for precision scheduling and monitoring," Comput. Electron. Agric., vol. 215, p. 108407, 2023.

[12] E. Pazouki, "A practical surface irrigation design based on fuzzy logic and meta-heuristic algorithms," Agric. Water Manag., vol. 256, p. 107069, 2021.

[13] S. Jaiswal and M. S. Ballal, "Fuzzy inference based irrigation controller for agricultural demand side management," Comput. Electron. Agric., vol. 175, p. 105537, 2020.

[14] R. S. Krishnan et al., "Fuzzy logic based smart irrigation system using Internet of Things," J. Clean. Prod., vol. 252, p. 119902, 2020.

[15] S. Jamroen et al., "An Intelligent irrigation scheduling system using low-cost wireless sensor network toward sustainable and precision agriculture," IEEE Access, vol. 8, pp. 172756–172769, 2020.

[16] W. R. Mendes et al., "Fuzzy control system for variable rate irrigation using remote sensing," Expert Syst. Appl., vol. 124, pp. 13–24, 2019.

[17] E. Bwambale et al., "Smart irrigation monitoring and control strategies for improving water use efficiency in precision agriculture: A review," Agric. Water Manag., vol. 260, p. 107324, 2022.

[18] O. Obaideen et al., "An overview of smart irrigation systems using IoT," Energy Nexus, vol. 5, p. 100124, 2022.

[19] G. Patrizi et al., "A Virtual Soil Moisture Sensor for Smart Farming Using Deep Learning," IEEE Trans. Instrum. Meas., vol. 71, pp. 1–11, 2022.

[20] N. Jihani et al., "Kalman filter based sensor fault detection in wireless sensor network for smart irrigation," Results Eng., vol. 20, p. 101395, 2023.

[21] H. Hamami et al., "Application of wireless sensor networks in the field of irrigation: a review," Comput. Electron. Agric., vol. 179, p. 105782, 2020.

[22] O. Friha et al., "Internet of Things for the future of smart agriculture: A comprehensive survey of emerging technologies," IEEE/CAA J. Autom. Sinica, vol. 8, no. 2, pp. 718–740, 2021.

[23] E. Giusti and S. Marsili-Libelli, "A Fuzzy Decision Support System for irrigation and water conservation in agriculture," Environ. Model. Softw., vol. 63, pp. 73–86, 2015.

[24] M. Li, R. Sui, Y. Meng, and H. Yan, "A real-time fuzzy decision support system for alfalfa irrigation," Comput. Electron. Agric., vol. 163, p. 104870, 2019.

[25] F. T. Teshome et al., "Improving soil moisture prediction with deep learning and machine learning models," Comput. Electron. Agric., vol. 226, p. 109414, 2024.

[26] F. Sadat Hashemi et al., "Integrating RS data with fuzzy decision systems for innovative crop water needs assessment," Int. J. Appl. Earth Observ. Geoinf., vol. 136, p. 104338, 2025.

[27] C. Catalano et al., "Anomaly detection in smart agriculture systems," Comput. Ind., vol. 143, p. 103750, 2022.

[28] M. Neethu et al., "Hybrid Transformer Network for Soil Moisture Estimation in Precision Irrigation," IEEE Access, 2024.

[29] S. Wu and J. Diao, "Outlier Detection in Wireless Sensor Networks: A Survey," IEEE Commun. Surveys Tuts., vol. 18, no. 1, pp. 589–623, 2016.

[30] M. Sami et al., "A deep learning-based sensor modeling for smart irrigation system," Agric. Water Manag., vol. 250, p. 106841, 2021.

[31] H. Hendra et al., "A smartphone-controlled, solar-powered automatic fertilizing and watering system for sustainable agriculture," Renew. Energy, vol. 198, pp. 123–132, 2023

https://doi.org/10.31436/iiumej.v27i2.3982

Steering Control Precision in Rack Steering Vehicles Using Finite-Time Prescribed Performance Anti-Windup PI Controller with Nonlinear Error Transformation

ADDIE IRAWAN[1*], NORSHARIMIE MAT ADAM[1], R.M.T. RAJA ISMAIL[1], ALIZA CHE AMRAN[2], MOHD ISKANDAR PUTRA AZAHAR[1]

[1]*Robotics, Intelligent Systems and Control Engineering (RiSC) Research Group, Faculty of Electrical & Electronics Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, Pahang, Malaysia.*
[2]*Motion Control Research Group (MCON), Fakulti Teknologi dan Kejuruteraan Elektrik (FTKE), Universiti Teknikal Malaysia Melaka (UTeM), Melaka, Malaysia.*

**Corresponding author: addieirawan@umpsa.edu.my*

(Received: 13 October 2025; Accepted: 22 January 2026; Published online: 10 May 2026)

ABSTRACT: This paper addresses the challenge of maintaining steering precision and yaw stability in rack steering vehicles (RSVs) operating under varying inertial loads and aerodynamic disturbances. Conventional control methods often struggle to ensure bounded tracking performance in the presence of actuator saturation and nonlinear dynamics. To overcome these limitations, a Finite-Time Prescribed Performance Anti-Windup PI Controller with Nonlinear Error Transformation (FPPC-API-NET) is proposed. The outer loop employs a finite-time prescribed performance framework with a nonlinear transformation to constrain the error trajectory within predefined decay envelopes, while the inner loop incorporates antiwindup compensation to mitigate actuator saturation effects. Simulation studies benchmarked against the FTPPC-API and API controllers demonstrate that FPPC-API-NET achieves up to a 95% reduction in steady-state error, a 70% reduction in settling time, and a 87% reduction in transient duration. The proposed method also enhances yaw stability, reduces lateral displacement by 35%, minimizes sideslip magnitude, and maintains tire-friction utilization below saturation, resulting in smoother lateral acceleration with up to 48% lower peaks. These results confirm that FPPC-API-NET offers superior robustness and precision during cornering maneuvers across varying inertia levels.

ABSTRAK: Kajian ini membincangkan cabaran dalam mengekalkan ketepatan kawalan stereng dan kestabilan sudut gerakan sisi bagi kenderaan jenis rak stering (RSV) yang beroperasi di bawah beban inersia berubah serta gangguan aerodinamik. Kaedah kawalan konvensional sering menghadapi kesukaran mengekalkan prestasi penjejakan apabila berlaku ketepuan aktuator dan kesan dinamik tak linear. Bagi mengatasi kekurangan ini, satu pengawal tirus berasaskan Finite-Time Prescribed Performance dengan Transformasi Ralat Tak Linear dan Antiwindup PI (FPPC-API-NET) telah dicadangkan. Gelung luar menggunakan rangka kerja prestasi terpiawai had masa dengan transformasi ralat tak linear bagi mengehadkan trajektori ralat dalam sempadan peluruhan yang ditetapkan, manakala gelung dalam aplikasi pampasan antiwindup bagi mengawal kesan ketepuan aktuator. Kajian simulasi berbanding pengawal FTPPC-API dan API menunjukkan bahawa FPPC-API-NET mencapai pengurangan ralat keadaan tunak sehingga 95%, masa penetapan 70% lebih pantas, dan tempoh peralihan 87% lebih singkat. Kaedah ini turut meningkatkan kestabilan gerakan sisi, mengurangkan anjakan lateral sebanyak 35%, meminimumkan gelinciran, serta mengekalkan penggunaan geseran tayar di bawah had ketepuan, sekali gus membuktikan

https://doi.org/10.31436/iiumej.v27i2.3982

keteguhan dan ketepatan unggul FPPC-API-NET dalam manuver membelok di bawah variasi inersia.

KEYWORDS: *Rack steering vehicle, prescribed performance control, nonlinear error transformation, inertial variation, cornering path.*

1. INTRODUCTION

Electric vehicles (EVs) and autonomous vehicles (AVs) introduce unique challenges in steering and lateral motion control. The instantaneous torque response and low center of gravity in EV platforms enhance maneuverability but can also exacerbate oversteering and yaw instability during aggressive cornering maneuvers. Meanwhile, AV systems must not only achieve accurate trajectory tracking but also operate reliably in the presence of sensor noise, actuation delays, and environmental uncertainties. These requirements demand steering control architectures that guarantee bounded tracking performance, robustness to inertial variations, and reliable operation under actuator constraints, particularly under high-curvature driving conditions.

To address these challenges, a wide range of vehicle control strategies has been investigated. Among the most established approaches are Model Predictive Control (MPC) [1], sliding mode control (SMC), and Takagi-Sugeno (T-S) fuzzy logic systems [2]. These frameworks have been further enhanced through the integration of intelligent techniques such as neural networks, reinforcement learning, and observer-based designs to accommodate uncertainties and time-varying dynamics [3]. Despite their effectiveness in simulation studies, many of these methods rely on simplified assumptions and often fail to adequately capture practical limitations such as actuator saturation, structural nonlinearities, and sensor bandwidth constraints, which can significantly degrade real-world performance.

Another limitation of many existing steering control approaches is their narrow focus on either stability or tracking performance, frequently overlooking robustness to disturbances, fault resilience, and implementation feasibility. Although integrated control schemes that combine steering, braking, and drive torque have been proposed, their practical deployment is often limited by high computational complexity and system integration challenges. Active Front Steering (AFS), a widely studied technique, improves steering responsiveness but is highly sensitive to inertial effects during sharp cornering, particularly in rack steering vehicle (RSV) systems [4]. Recent advances have introduced coordinated ESP+AFS controllers using variable sliding models to enhance lateral stability under low-adhesion conditions [5]. Comparative studies among AFS, front-wheel independent steering (FWIS), four-wheel steering (4WS), and four-wheel independent steering (4WIS) further indicate that FWIS and 4WIS architectures outperform conventional AFS in yaw rate regulation and path-tracking performance under nonlinear dynamics [6].

Managing inertia-induced instability during cornering remains a persistent and underexplored challenge in RSV dynamics. Rapid steering inputs generate large lateral loads that can lead to vehicle sideslips, loss of tracking accuracy, and, in extreme cases, complete destabilization. Traditional control strategies are often insufficient to mitigate these inertial effects. Rsetam et al. suggested that impedance-based control techniques can improve steering response by modulating control input according to environmental interaction, making them suitable for systems affected by aerodynamic drag and road surface variation [7].

Another promising direction involves the application of Super-Twisting Sliding Mode Control (ST-SMC), which decouples lateral and longitudinal dynamics to enhance robustness

to inertia-driven disturbances [8]. Integrated control strategies combining AFS with direct yaw moment control (DYC) have demonstrated improved lateral stabilization performance [9], while inertial delay control-based sliding mode controllers for 4WS systems with variable steering ratios have been shown to enhance yaw stability under uncertain tire cornering stiffness. In addition, robust output-feedback control designs for AFS have been proposed to mitigate side-wind disturbances and parameter uncertainties, thereby improving lateral stability margins [10].

In parallel, Prescribed Performance Control (PPC) has gained increasing attention as a systematic framework for guaranteeing system performance within predefined error boundaries, even in the presence of disturbances and uncertainties. Finite-time PPC strategies have been successfully applied to steer-by-wire systems [11], while fault-tolerant PPC schemes incorporating fuzzy logic have demonstrated improved reliability under actuator faults [12]. Gao et al. proposed a finite-time PPC strategy for vehicular platoons that ensures bounded tracking and spacing errors [13], whereas Nguyen et al. developed an event-triggered PPC framework with disturbance observers to address electro-hydraulic actuation delays [14].

Further refinements have been achieved by incorporating high-gain observers, adaptive sliding-mode control, and dynamic threshold strategies to enhance tracking performance [13]. Despite these advances, Proportional–Integral–Derivative (PID) controllers remain widely adopted in steering systems due to their simplicity and ease of implementation. However, conventional PID controllers often lack robustness in handling actuator saturation and highly nonlinear vehicle behavior. Anti-windup compensation techniques have been introduced to alleviate integrator saturation, while vector-field methods and intelligent optimization have been employed to enhance PID performance in various steering and motion control applications [15,16].

Therefore, this study addresses the control challenges associated with steering precision and yaw regulation in rack-steering vehicles, particularly under high-curvature cornering conditions, where oversteering and inertia-induced instability are most prevalent. A novel control framework is proposed that integrates Finite-Time Prescribed Performance Control (FTPPC), based on [17], with an Anti-Windup Proportional–Integral (AW-PI) structure and a bounded Nonlinear Error Transformation (NET). Unlike existing hybrid MPC- or SMC-based approaches that often require model inversion, discontinuous control actions, or high computational effort, the proposed FPPC-API-NET achieves finite-time convergence using smooth, bounded transformations while explicitly addressing actuator saturation. Moreover, this proposed control framework is evaluated through comprehensive numerical simulations using a nonlinear RSV dynamic model, thereby enabling a systematic assessment of steering precision and yaw stability under variations in inertia and aerodynamic disturbances. Owing to its algebraic nonlinear transformations and PI-based structure, the proposed controller is computationally lightweight and suitable for real-time embedded automotive implementation, although experimental validation is beyond the scope of this study.

The remainder of this paper is structured as follows. Section 2 presents the dynamic modeling of the RSV, including the Dugoff tire formulation and the coupled yaw–lateral–roll motion equations. Section 3 describes the formulation of the proposed FPPC-API-NET, detailing the control architecture, constraint handling, and nonlinear transformation laws. Section 4 presents simulation results and a comparative performance analysis of FPPC-API-NET, FTPPC-API, and API controllers under varying inertial and aerodynamic conditions. Finally, Section 5 summarizes the key findings and outlines directions for future research.

https://doi.org/10.31436/iiumej.v27i2.3982

2. CONTROLLER DESIGN

This study adopts a cascaded two-loop steering-position controller for RSV's steering-angle control. The outer loop implements an adaptive FT-PPC to target the transformation of the steering angular error, while the inner loop employs an Anti-windup PI (AW-PI) controller to regulate the steering angular position based on the transformed errors. The proposed cascaded controller enforces error constraints using a tunable prescribed performance function, while the anti-windup function ensures robustness against saturation.

2.1. Overview of Steering Vehicle Dynamics

The RSV is modeled to capture the essential dynamics governing interactions among steering, yaw, and roll. The model employs the Dugoff tire formulation to describe nonlinear longitudinal and lateral tire forces, while the vehicle states are defined by the sideslip angle, β, yaw rate, $\dot{\psi}$, and roll angle, ϕ . The overall system coordination, including the vehicle motion, reference axes, and tire dynamics, is depicted in Fig. 1.

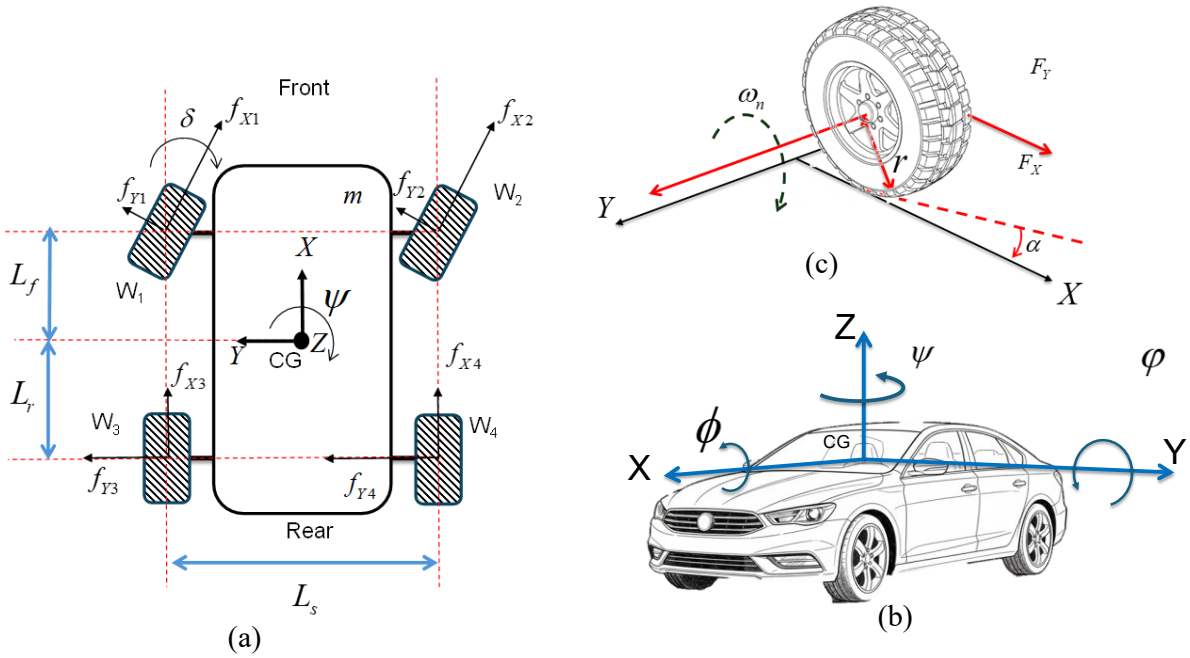


Figure 1. RSV system coordination, (a) vehicle dynamics motion, (b) vehicle axis system, and (c) tire dynamics motion.

The kinematic accelerations in the inertial frame are expressed as follows:

$$a_x = \dot{V}_x - \dot{\psi}V_y, \quad a_y = \dot{V}_y - \dot{\psi}V_x \tag{1}$$

Then, the slip angles of the front and rear tires, which directly influence the lateral forces, are given by

$$a_f = \delta_f - \arctan\left(\frac{V_y + L_f\dot{\psi}}{V_x}\right), \quad a_r = -\arctan\left(\frac{V_y - L_r\dot{\psi}}{V_x}\right) \tag{2}$$

where δ_r is the steering input, L_f and L_r are the distances from the vehicle center of gravity (CG) to the front and rear axles, respectively. The Dugoff tire model [18] provides the longitudinal and lateral forces as:

$$F_x = C_\sigma \sigma_x f(\lambda), \quad F_y = C_\alpha \tan\alpha f(\lambda) \tag{3}$$

https://doi.org/10.31436/iiumej.v27i2.3982

where C_σ and C_α are longitudinal and cornering stiffness coefficients, respectively, σ_x is the longitudinal slip ratio, and $f(\lambda)$ is a nonlinear saturation function dependent on the road–tire friction coefficient, μ. The parameters used in this RSV dynamic model are summarized in Table 1 and are adopted from established Dugoff tire–based vehicle models in the [18,19]. The listed parameters represent the physical properties, stiffness coefficients, geometric dimensions, and inertial constants governing the vehicle's behavior under steering motion.

The longitudinal slip ratio is defined as follows:

$$\sigma_x = \begin{cases} \frac{r\omega - V_x}{r\omega}, & acceleration \\ \frac{V_x - r\omega}{V_x}, & braking \end{cases} \tag{4}$$

with r denoting the tire radius and ω is the wheel's angular velocity and $n = 1,2,\ldots,4$ is the number of wheels as depicted in Fig.1(c). The yaw moment balance around the vertical axis is expressed as follows:

$$I_x\ddot{\psi} + I_{xz}\ddot{\phi}_v = \sum M_{z,i} \tag{5}$$

where I_z is the yaw inertia and I_{xz} is the cross-inertia coupling between yaw and roll. The roll dynamics are described by

$$I_{xx}\ddot{\phi} + C_\phi\dot{\phi}_v + K_\phi\phi_v = m_s h a_y \tag{6}$$

where I_{xx} is roll inertia, C_ϕ and K_ϕ are the suspension damping and stiffness coefficients, m_s is the spring-mass, and h, is the height of the center of gravity.

Table 1. List of RSV's parameters and coefficients

Definition	Symbol	Value
Vehicle mass	m	1300 kg
Longitudinal stiffness	C_σ	20000 N/rad
Cornering stiffness	C_α	80000 N/rad
Distance of front axle from CoG	L_f	1.180 m
Distance of rear axle from CoG	L_r	1.770 m
Track width between left and right wheels	L_s	1.880 m
Acceleration due to gravity	g	9.81 m/s²
Tire radius (effective)	r	0.38m
Tire–road friction coefficient	μ	0.85
Yaw moment of inertia	I_z	2480 kg·m²

By applying small-angle approximations to the slip angles and assuming constant longitudinal velocity, the system reduces to the well-known bicycle model, expressed as

$$\dot{\beta} = -\left(\frac{C_f + C_r}{mV_x}\right)\beta + \left(\frac{C_r L_r - C_f L_f}{mV_x^2} - 1\right)\dot{\psi} + \left(\frac{C_f}{mV_x}\right)\delta_f \tag{7}$$

$$\ddot{\psi} = \left(\frac{C_f L_f - C_r L_r}{I_x}\right)\beta - \left(\frac{C_f L_f^2 - C_2 L_r^2}{I_z V_z}\right)\dot{\psi} + \left(\frac{L_f C_f}{I_z}\right)\delta_f \tag{8}$$

where C_f and $C_r$denote the front and rear cornering stiffness, respectively, and m, is the vehicle mass. The complete RSV dynamic model structure, incorporating the Dugoff tire formulation and the lateral–yaw–roll interaction, is illustrated in Fig. 2. This figure highlights the integration of subsystems such as tire dynamics, yaw and roll coupling, and the vehicle motion response, which collectively represent the plant used in the proposed control design.

https://doi.org/10.31436/iiumej.v27i2.3982

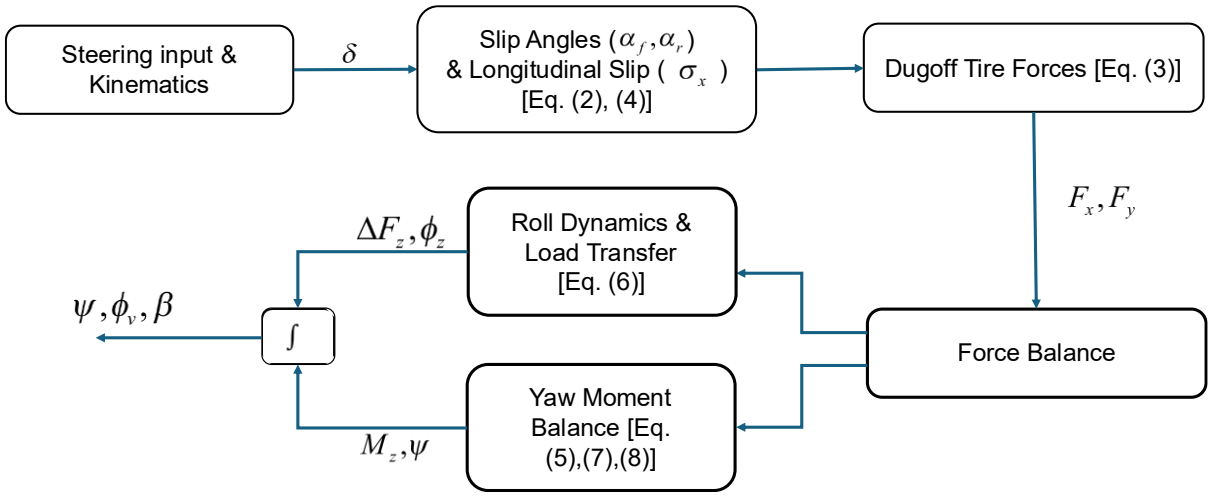


Figure 2. Overall structure of the RSV dynamics model

2.2. Problem Formulations

Let the yaw angle output be ψ and reference be ψ_r. Define yaw tracking error as follows:

$$e_\psi(t) = \psi_r(t) - \psi(t) \tag{9}$$

Adopt the prescribed performance function (PPF) [17] as follows:

$$\rho(t) = (\rho_0 - \rho_\infty)e^{-\frac{t^2}{2\left(\frac{t_c}{2}\right)^2}} + \rho_\infty, \ \ \rho_0 > \rho_\infty > 0, t_c > 0, \tag{10}$$

and form the normalized and transformed errors as follows:

$$\Delta(t) = \frac{e_\psi}{\rho(t)}, \quad \varepsilon(t) = \frac{1}{2}\ln\left(\frac{1+\Delta(t)}{1-\Delta(t)}\right) \tag{11}$$

Then, differentiation gives

$$\dot{\varepsilon}(t) = \frac{1}{1-\Delta^2(t)}\left(\frac{\dot{e}_\psi(t)}{\rho(t)} - \Delta(t)\frac{\dot{\rho}(t)}{\rho(t)}\right) \tag{12}$$

where $\dot{e}_\psi$ is the yaw error rate that can be expressed as follows:

$$\dot{e}_\psi(t) = \dot{\psi}_r(t) - \dot{\psi}(t) \tag{13}$$

with $\dot{\psi}_r$ is the reference for the vehicle's yaw rate.

2.3. Finite-Time Prescribed Performance Anti-Windup PI Controller with Nonlinear Error Transformation

Starting from the prescribed performance function, $\rho(t)$ and the transformed error, $\varepsilon(t)$, defined in Eqs. (10)-(12), the control objective is to ensure that the yaw tracking error remains strictly within the prescribed performance bounds while generating a steering command that respects actuator limitations. In practical rack steering systems, direct inversion of the steering–yaw dynamics, or aggressive high-gain control, may lead to actuator saturation, excessive control effort, or degraded robustness. Therefore, instead of explicitly inverting the system dynamics, the proposed approach shapes the steering correction as a smooth nonlinear function of the transformed error, enabling finite-time convergence while preserving bounded control authority.

https://doi.org/10.31436/iiumej.v27i2.3982

Conventional prescribed performance control schemes commonly employ logarithmic or unbounded power-based error transformations to enforce performance constraints [17]. Although effective in theory, such transformations may produce excessively large control signals when the transformed error grows, particularly during large transients or under inertia variation. This behavior is undesirable in rack-steering vehicles, where physical steering limits and actuator saturation must be explicitly accounted for. To address this limitation, a bounded Nonlinear Error Transformation (NET) is introduced in this study that preserves the sign information of the transformed error while ensuring smooth, bounded control actions.

Specifically, a sign-preserving power of the transformed error is defined as

$$\varepsilon^{\gamma}(t) = sign\big(\varepsilon(t)\big)|\varepsilon(t)|^{\gamma}, \;\; 0 < \gamma \leq 1 \tag{14}$$

which retains sensitivity to small errors while moderating the growth rate for large error magnitudes. This transformed error is then mapped to the steering domain through a bounded and monotonic nonlinear function given by

$$\delta_{fb}(t) = \Delta_{max} \tanh\left(\frac{K_{NET}\varepsilon^{\gamma}(t)}{\Delta_{max}}\right), \qquad K_{NET} > 0, \Delta_{max} > 0 \tag{15}$$

where $\delta_{fb}(t)$ denotes the feedback steering correction, K_{NET} is a positive gain that scales the response near the origin, and Δ_{max} defines the maximum allowable steering correction. The hyperbolic tangent function ensures that the feedback term is odd, strictly increasing, and bounded such that $|\delta_{fb}| \leq \Delta_{max}$, thereby preventing excessive steering commands and preserving actuator safety. For small values of $|\varepsilon(t)|$, the nonlinear function in Eq. (15) can be locally approximated using the first-order Taylor expansion of the hyperbolic tangent function, yielding

$$\delta_{fb}(t) \approx \Delta_{max} K_{NET}\varepsilon^{\gamma}(t) \tag{16}$$

which indicates that K_{NET} governs the proportional response in the vicinity of the equilibrium point. The exponent $\gamma \in (0,1)$ further adjusts the nonlinearity of the transformation, where smaller values of γ enhance sensitivity for small errors and accelerate convergence toward the prescribed performance funnel, while $\gamma = 1$ recovers a linear behavior.

For large values of $|\varepsilon(t)|$, the hyperbolic tangent function smoothly saturates, causing $\delta_{fb}(t)$ to asymptotically approach $\pm\Delta_{max}$. This bounded behavior fundamentally distinguishes the proposed NET from conventional logarithmic or unbounded power-based transformations, as it prevents aggressive transients from demanding excessive steering authority and improves robustness under actuator saturation.

To ensure a finite slope at the origin when $\gamma < 1$, a regularized form of the power function may be employed

$$\varepsilon_{\gamma,\delta}(t) = sign\big(\varepsilon(t)\big)[|\varepsilon(t)| + \delta]^{\gamma} \tag{17}$$

which preserves the oddness and boundedness of the transformation while avoiding singular behavior. In this case, $\varepsilon^{\gamma}(t)$ in Eq. (16) is replaced by $\varepsilon_{\gamma,\delta}(t)$. The steering reference is then constructed as the sum of a feedforward term and the bounded feedback component,

$$\delta_v(t) = \delta_r(t) + \delta_{fb}(t) \tag{18}$$

where $\delta_v(t)$ represents a new or dynamically updated desired steering trajectory. Thus, the steering tracking can be expressed as follows, with $\delta(t)$ as a measured feedback steering from the RSV dynamic model [20]:

https://doi.org/10.31436/iiumej.v27i2.3982

$$e_\delta(t) = \delta_v(t) - \delta(t). \tag{19}$$

The inner-loop steering controller employs an anti-windup PI (AW-PI) structure with back-calculation, expressed as

$$\begin{aligned} u_c(t) &= K_p e_\delta(t) + K_i z(t) \\ \dot{z}(t) &= e_\delta(t) + K_b(u(t) - u_c(t)) \\ u(t) &= \Lambda(u_c(t)) \end{aligned} \tag{20}$$

Here, $z(t)$ is the integral state, $K_b > 0$ is the back calculation gain, and $\Lambda(\cdot)$ denotes the actuator map, which may impose magnitude or rate constraints. In the linear operating region, $\Lambda(u_c) = u_c$ and the integrator evolves with $\dot{z} = e_\delta$. When the actuator saturates, the term, $K_b(u - u_c)$, drives the integral state toward the actual actuator signal, which prevents windup and ensures smooth recovery as the actuator reenters its linear range [20]. Finally, $[K_p, K_i] > 0$ are the positive finite design parameters. With the inner loop tuned to track, δ_v, faster than the outer loop, the NET in Eqs. (14)-(18) produces a steering correction that is sign consistent with ε, increases with $|\varepsilon|$, yet remains bounded by Δ_{max}, therefore, the prescribed bound, $|e_\psi| < \rho(t)$ is preserved while respecting steering limits. Fig. 3 shows the overall proposed control system structure with the RSV dynamic model as a plant.

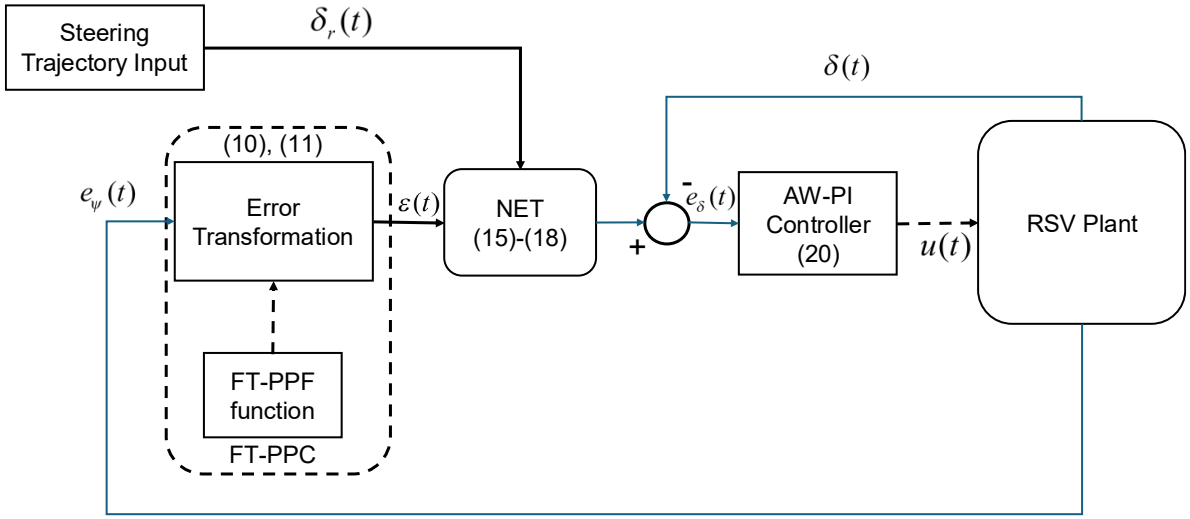


Figure 2. Functional block of FPPC-API-NET for RSV's steering positioning

2.4. Stability Analysis

Consider the $\varepsilon(t)$defined in Eq. (11), which relates to the yaw tracking error, $e_\psi(t)$under the prescribed performance transformation. To verify the stability of the cascaded control framework, define the Lyapunov candidate as

$$V_\varepsilon = \frac{1}{2}\varepsilon^2(t), \tag{21}$$

which is positive definite for all $\varepsilon(t) \neq 0$. Its time derivative along the trajectories of the closed-loop error system is obtained as

$$\dot{V}_\varepsilon = \varepsilon(t)\dot{\varepsilon}(t). \tag{22}$$

From Eq.(12), the derivative of $\varepsilon(t)$can be expressed as

$$\dot{\varepsilon}(t) = \frac{2\dot{e}_\psi(t)}{\rho(t)\left[1-\left(e_\psi(t)/\rho(t)\right)^2\right]} - \frac{2e_\psi(t)\dot{\rho}(t)}{\rho^2(t)\left[1-\left(e_\psi(t)/\rho(t)\right)^2\right]}. \tag{23}$$

Under the outer-loop FT-PPC NET control law Eqs. (15)–(18), the transformed error dynamics can be shaped as

$$\dot{\varepsilon}(t) = -k_\varepsilon(t)\varepsilon(t) + \Delta(t), \tag{24}$$

with $k_\varepsilon > 0$ denotes the adaptive convergence rate determined by the prescribed performance shaping, and $\Delta(t)$ represents bounded residual effects arising from model uncertainties or coupling terms in the RSV dynamics. Substituting this into the Lyapunov derivative yields

$$\dot{V}_\varepsilon(t) = -k_\varepsilon(t)\varepsilon^2(t) + \varepsilon(t)\Delta(t). \tag{25}$$

Since $\Delta(t)$ is bounded, there exists a positive constant, $\gamma > 0$ such that

$$\dot{V}_\varepsilon(t) \leq -\gamma V_\varepsilon(t), \tag{26}$$

which guarantees exponential convergence of $\varepsilon(t)$ to zero. Consequently, the original $e_\delta(t)$, is confined within the prescribed performance envelope, $\rho(t)$ defined in Eq. (10), ensuring finite-time boundedness and overshoot-free tracking. For the inner loop, the AW-PI control law Eq. (20) regulates the steering error, $e_\delta(t)$ with control input, $u_c(t)$ and integrator dynamics, $\dot{z}(t)$. The Lyapunov function is selected as

$$V_\delta(t) = \frac{1}{2}e_\delta^2(t) + \frac{1}{2K_i}z^2(t), \tag{27}$$

whose time derivative along Eq. (20) satisfies

$$V_\delta(t) \leq -K_p e_\delta^2(t) - K_b[u_c(t) - \Lambda(u_c(t))]^2 \leq 0. \tag{28}$$

Hence, $e_\delta(t)$ and $z(t)$ are bounded and converge asymptotically to zero when the actuator operates within its linear range, while the anti-windup term ensures smooth recovery under saturation. Since both subsystems are input-to-state stable (ISS) and the inner loop is designed to respond faster, $T_\delta \ll T_\varepsilon$, the small-gain theorem ensures the stability of the full cascaded system. Therefore, all closed-loop signals are bounded, and the tracking objectives are achieved.

The proposed FPPC-API-NET controller ensures that the transformed error, $\varepsilon(t)$, converges exponentially to zero, maintaining $e_\psi(t)$ strictly within the prescribed bounds. The inner loop guarantees bounded steering control with effective anti-windup compensation. The overall system achieves robust stability and tracking performance in the presence of model uncertainties, actuator saturation, and external disturbances.

3. Results and Discussion

Simulation studies were conducted to assess and verify the proposed FPPC-API-NET controller's ability to regulate steering precision and yaw stability under nonlinear coupling effects. Prior to simulation, the FPPC-API-NET parameters were fine-tuned to satisfy the nonlinear and coupling constraints embedded in its control structure, with both the outer and inner control loops optimized to maintain consistency with the RSV model's nonlinear dynamics. Specifically, the outer-loop finite-time prescribed performance NET was calibrated to ensure strict error confinement within the decaying envelope defined by the nonlinear transformation function, $\rho(t)$, while maintaining smooth convergence under varying inertial and aerodynamic disturbances. Concurrently, the inner anti-windup PI (AW-PI) parameters were tuned via constraint-aware iterations to satisfy actuator saturation limits and the coupling between the steering torque dynamics and the yaw-moment balance described in Eqs. (5)-(8). This parameter adjustment ensured that the nonlinear saturation characteristics in the NET law (Eqs. 15–16) and the feedback regularization in the AW-PI loop (Eq. 20) operated in a

https://doi.org/10.31436/iiumej.v27i2.3982

dynamically coordinated manner, preserving stability and maintaining bounded steering authority within the prescribed performance envelope. The finalized parameter set, summarized in Table 2, represents the optimum balance between transient responsiveness and saturation avoidance, determined through iterative closed-loop validation using the RSV plant. To emulate real on-road conditions, aerodynamic disturbances were introduced into the RSV model within the 9–10 Hz frequency band, along with an additional 100 kg payload to represent inertia-variation effects. The comparative analysis was carried out by benchmarking the proposed controller against the Finite-Time Prescribed Performance with Antiwindup PI (FTPPC-API) and the baseline Antiwindup PI (API) controllers used in previous studies[20].

The steering position tracking responses in Fig. 4 indicate that the proposed FPPC-API-NET achieves the highest closed-loop tracking accuracy and the fastest convergence under varying inertia conditions. The response closely follows the steering reference, exhibits negligible steady-state error, and shows rapid transient decay, with settling occurring at approximately 2 seconds despite the imposed disturbances. The FTPPC-API yields moderate performance with a larger residual steady-state deviation and a longer settling time, while the baseline API exhibits pronounced lag and oscillatory behavior. Quantitatively, FPPC-API-NET reduces the final steady-state error by approximately 68% relative to FPPC-API and nearly 95% relative to API, confirming superior steering-angle regulation and transient suppression.

This improvement is primarily attributed to the decay-function envelope of the prescribed performance function (PPF), as shown in Fig. 3, which dynamically constrains the evolution of the tracking error within predefined boundaries. The smooth funnel-shaped envelope defined by $\rho(t)$ effectively shapes the transient response and limits overshoot while enforcing finite-time convergence.

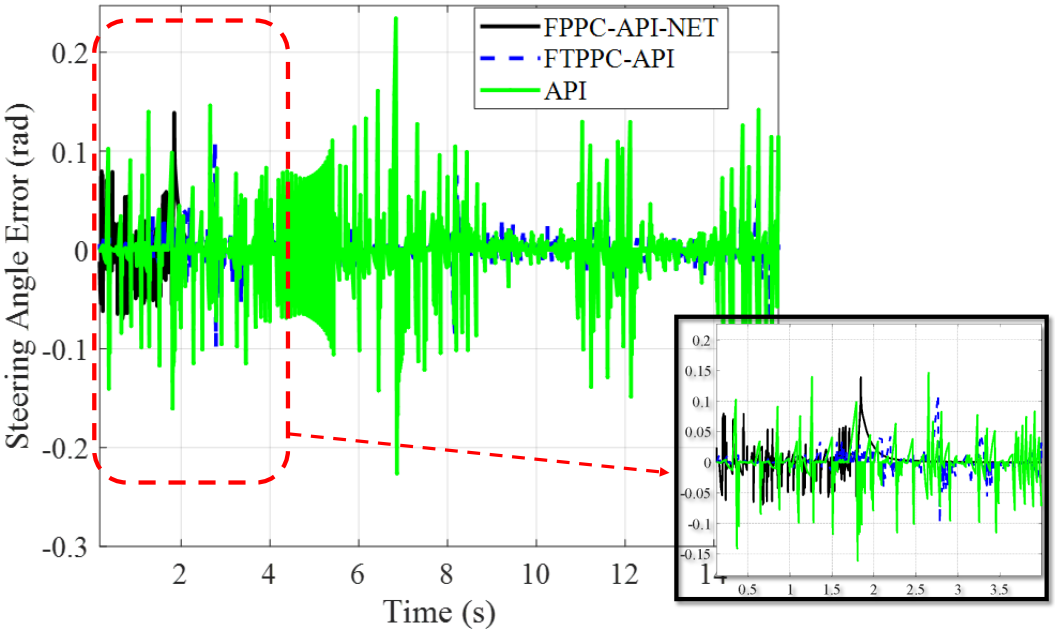


Figure 3. Steering position tracking performances between controllers.

In the FPPC-API-NET, the nonlinear error transformation (NET) acts as an adaptive gain-scaling mechanism that regulates the control effort based on the magnitude of the transformed error, ensuring that large initial deviations produce proportionally stronger corrective actions while small residual errors are attenuated smoothly near convergence. Consequently, the tracking-error trajectory remains strictly confined within the prescribed decay limits throughout the maneuver.

https://doi.org/10.31436/iiumej.v27i2.3982

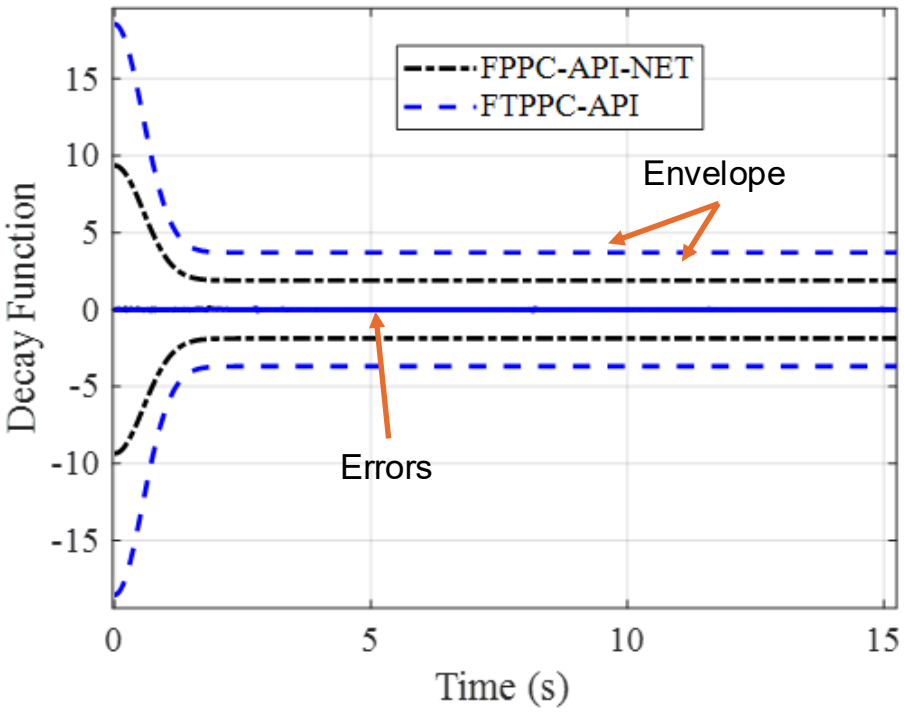


Figure 4. Sample of tracking error in the decay smooth function boundary between FPPC-API-NET and FTPPC-API.

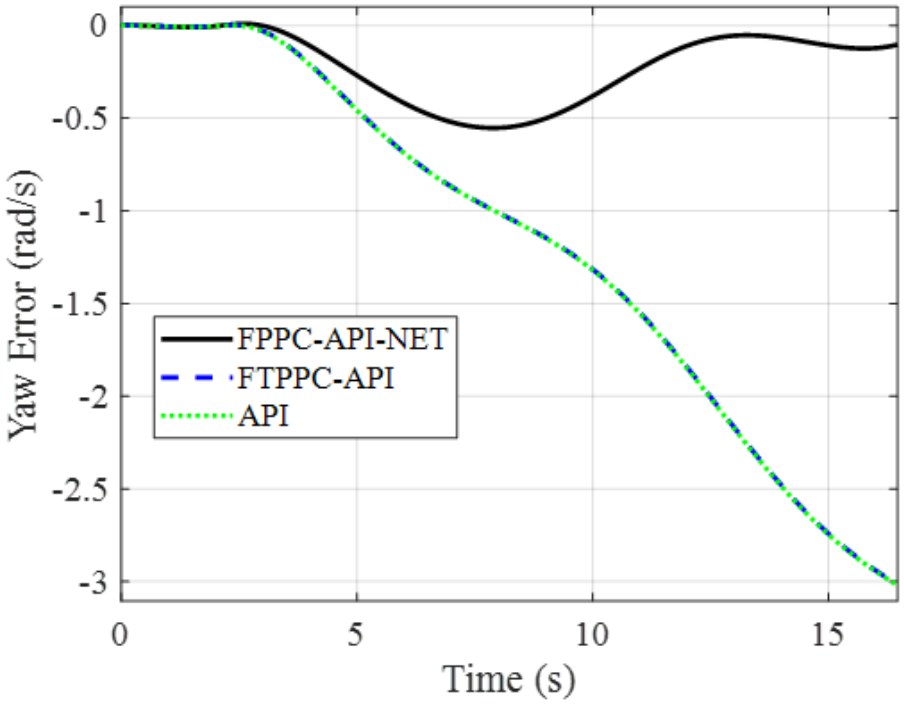


Figure 5. Comparison between controllers in the vehicle's yaw stability margin

Both FPPC-API-NET and FTPPC-API maintain their error trajectories within the respective PPF envelopes. However, the proposed FPPC-API-NET exhibits a narrower performance funnel, approximately 23 % tighter, which imposes stricter error regulation and enhances the convergence rate. The nonlinear transformation further introduces a nonlinear damping effect that improves transient-response uniformity, preventing high-frequency oscillations and reducing the control energy required for error correction. This explains the

https://doi.org/10.31436/iiumej.v27i2.3982

faster convergence and smaller steady-state deviation observed in Fig. 4. Overall, the enhanced steering precision achieved by FPPC-API-NET stems directly from the combined effect of finite-time PPF shaping and bounded NET, which collectively improve control adaptability under varying inertial conditions.

These enhanced steering characteristics directly influence the vehicle's yaw motion, as shown in Fig. 6. FPPC-API-NET maintains stable yaw-rate responses with uniform gain and phase margins across inertia variations, indicating robustness to dynamic parameter changes. The FTPPC-API response, by contrast, exhibits a slower transient response as inertia increases, while the API controller suffers from pronounced overshoot and phase lag, indicating inferior damping and poor stability. FPPC-API-NET's improved damping characteristics mitigate oscillations and preserve bandwidth, confirming that the proposed NET effectively compensates for inertia-induced dynamics, thereby sustaining yaw stability throughout cornering maneuvers.

The effect of improved yaw regulation is further reflected in lateral-displacement performance, presented in Fig. 6. FPPC-API-NET achieves the smallest deviation from the reference path, reducing peak lateral displacement by approximately 35 % compared with FTPPC-API and more than 60 % relative to API. Moreover, FPPC-API-NET exhibits a faster lateral-error decay rate (≈ 40 % improvement), ensuring rapid trajectory recovery following a disturbance. These results highlight the controller's superior capability to manage lateral–yaw coupling through adaptive force redistribution between the front and rear axles.

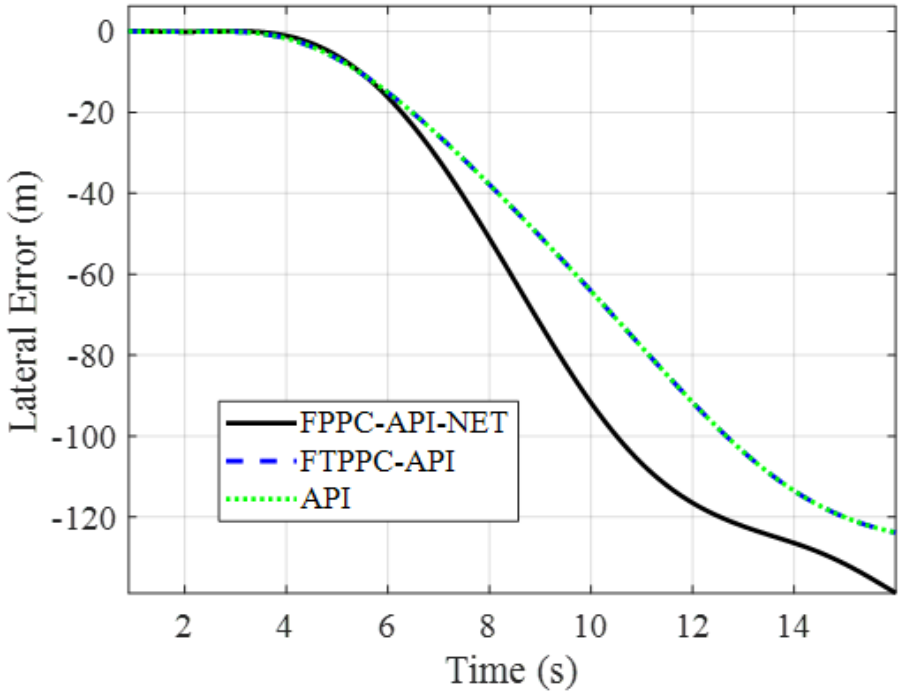


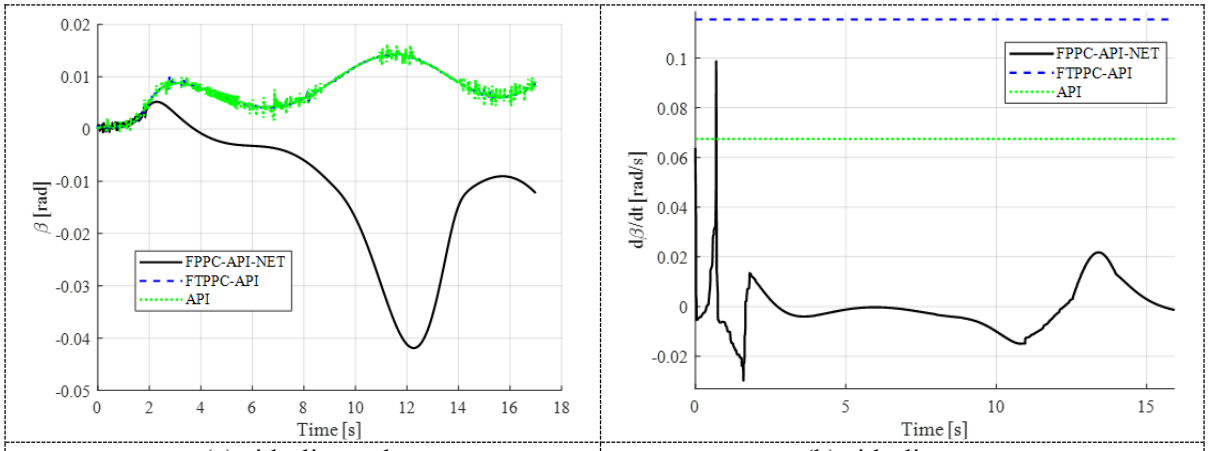
Figure 6. Comparison of the vehicle's lateral displacement errors between controllers.

The tighter integration of the outer PPF layer and the inner anti-windup compensation ensures that transient steering corrections remain dynamically consistent with tire-force limits, maintaining both path-tracking accuracy and yaw stability even under inertia-driven nonlinearities.

Table 2. Fine-tuned Design Parameters

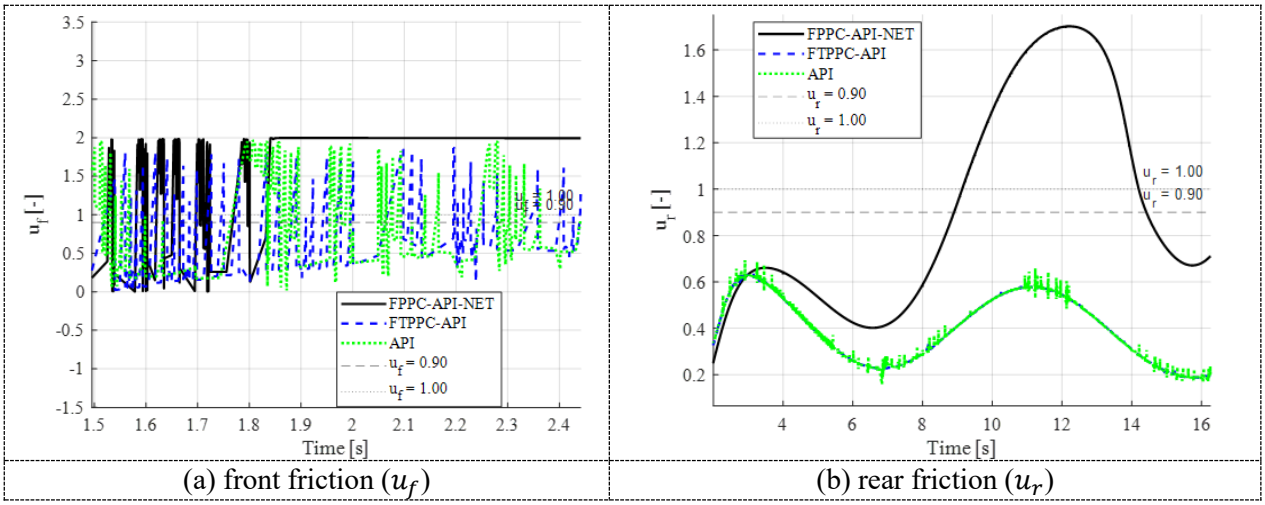
Parameter	FPPC-API-NET	FTPPC-API	API
K_p	15.5017	11.3057	14.0137
K_i	8.7521	1.3935	4.2017
K_b	3.5911	7.5534	0.1
ρ_0	9.3491	16.7776	-
t_c	2.2289	1.2006	-
ρ_∞	1.8855	4.6328	-
K_{NET}	0.0689	-	-
Δ_{max}	0.0478	-	-
γ	0.5360	-	-

Complementary insights are provided by the sideslip dynamics in Figs. 7(a) and 7(b). FPPC-API-NET maintains the lowest sideslip angle, β, magnitude, and a faster sideslip rate, $d\beta/dt$, decay. The settling time of $d\beta/dt$ is approximately 45% shorter than that of FTPPC-API and 70% shorter than API, illustrating enhanced lateral stability and superior disturbance rejection capability. Lower βamplitudes also correspond to improved tire force alignment and reduced lateral drift, confirming the controller's effectiveness in maintaining balanced cornering behavior.



(a) sideslip angle (b) sidesling rate

Figure 7. Comparison between controllers



(a) front friction (u_f) (b) rear friction (u_r)

Figure 8. Comparison between the vehicle's front and rear friction utilization.

Further evidence of robustness is provided by the tire-friction utilization results in Figs. 8(a) and 8(b). Both front and rear tires operate within the safe adhesion limit of $\mu = 0.85$, demonstrating that FPPC-API-NET prevents tire saturation during high-curvature maneuvers. The front-tire utilization peak is approximately 0.91 for FPPC-API-NET, compared with 0.96 for FTPPC-API and 1.04 for API. The rear-tire utilization remains near 0.88, indicating balanced load transfer between axles. These findings confirm that the nonlinear error-transformation control law not only enhances steering precision but also maintains efficient friction distribution, resulting in improved cornering stability, lower energy loss, and extended tire lifespan. Although explicit electrical energy consumption is not directly evaluated in this study, the observed reduction in steering command oscillations, lower tire–road friction utilization, and smoother lateral acceleration profiles collectively serve as indirect indicators of improved control efficiency. By preventing excessive corrective steering actions and avoiding tire force saturation during aggressive cornering maneuvers, the proposed FPPC-API-NET reduces unnecessary energy dissipation while maintaining robust lateral stability. These characteristics suggest that the proposed control strategy is well-suited to energy-conscious steering applications, particularly on electric and autonomous vehicle platforms, where efficiency and stability must be considered jointly.

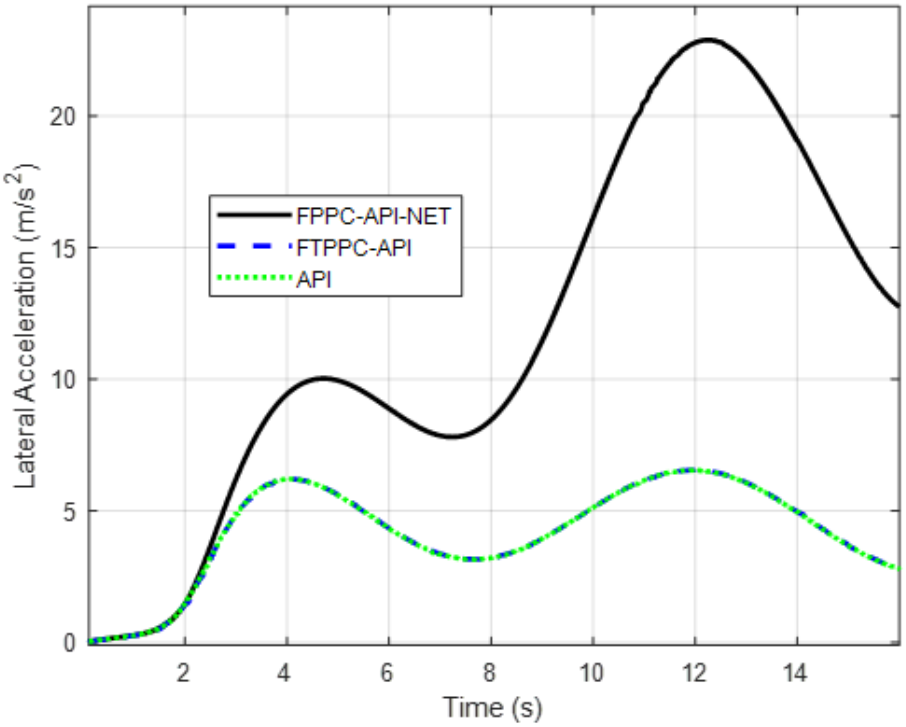


Figure 9. Comparative lateral acceleration responses (a_y), among controllers

This efficiency is also evident in the lateral acceleration, a_y, responses plotted in Fig. 9. FPPC-API-NET delivers smoother acceleration profiles with reduced oscillations and peak magnitudes, achieving approximately 28% lower peak magnitudes than FTPPC-API and 48% lower than API. The smoother $a_y$curves indicate improved ride comfort, reduced dynamic load transfer, and minimized roll excitation, all of which contribute to enhanced handling performance and passenger comfort under varying inertial conditions.

The observed dynamic responses are consistent with the robustness indices summarized in Table 3, which compare transient specifications across controllers. FPPC-API-NET demonstrates a 66.7% faster rise time, a 87.4% shorter transient duration, and a 70% faster

settling time than FPPC-API, while maintaining stable bounded overshoot. Although the overshoot amplitude is approximately 221.8% larger than that of FTPPC-API, it is well-regulated and quickly attenuated within the prescribed performance boundary, reflecting the controller's aggressive yet stable finite-time convergence. The combined results affirm that FPPC-API-NET offers the best trade-off between fast convergence and robust stability across steering and yaw dynamics.

Table 3. Performance and robustness of the controllers

Controller / Parameter	FPPC-API-NET	FTPPC-API	API
Rise Time	0.1 ms	0.3 ms	0.5 ms
Transient Time	2.14 ms	8.18 ms	10.1 ms
Settling Time	2.90 ms	8.18 ms	9.8 ms
Max. Overshoot	6.58 milliradians	1.09 milliradians	460.82 milliradians

Table 4. Overall measurement matrix of controller performance

Metric	FPPC-API-NET	FTPPC-API	API	Improvement (FPPC-API-NET)
Steering steady-state error	Lowest (≈95% ↓ vs API)	Moderate	Highest	95% lower
Tracking error bound	Tightest (23% ↓)	Moderate	Not bounded	23% smaller
Yaw stability margin	High, uniform	Medium, varying	Low	Consistent stability
Lateral displacement peak error	−35%	Moderate	Highest	35% lower
Sideslip angle amplitude	Lowest	Medium	Highest	45–70% faster decay
Tire friction (front/rear)	0.91 / 0.88	0.96 / 0.94	>1.00 / >1.00	Avoids saturation
Lateral acceleration peak	28% lower	Moderate	Highest	48% lower vs API
Control energy	Lowest	Medium	Highest	Most efficient

The integrated comparison is listed in Table 4, which summarizes the performance measurement matrix across all evaluated criteria. The FPPC-API-NET demonstrates consistently superior results, achieving higher steering accuracy, stronger yaw stability, reduced sideslip, more balanced tire utilization, smoother lateral acceleration, and greater overall robustness. In contrast, the FTPPC-API delivers moderate performance with acceptable tracking and stability, whereas the API controller shows the weakest control quality, characterized by larger errors, slower convergence, and less effective dynamic regulation.

4. CONCLUSIONS

This study presents a proposed FPPC-API-NET for steering and yaw-motion regulation of RSV under varying inertial loads and aerodynamic disturbances. By integrating an FT-PPC framework with NET, the proposed controller ensures that tracking errors remain bounded within predefined decay envelopes, while the anti-windup compensation maintains actuator stability. Simulation results verify that FPPC-API-NET in RSV drive motion in the cornering path achieves the most accurate steering position tracking, reducing steady-state error by 68% compared to FTPPC-API and 95% relative to API, with faster convergence and minimal

overshoot. These improvements directly enhance yaw stability and path-following accuracy, yielding smaller lateral displacement errors, smoother yaw responses, and better overall dynamic coordination during cornering maneuvers.

The improved steering and yaw performance further enhance vehicle stability, as evidenced by reduced sideslip magnitude and rate, balanced tire friction utilization below saturation limits, and smoother lateral acceleration, with peaks up to 48% lower than the baseline. Robustness analysis confirms that FPPC-API-NET exhibits a 66.7% faster rise time, a 87.4% shorter transient duration, and a 70% faster settling time than FPPC-API, while maintaining tight prescribed performance bounds. These results establish FPPC-API-NET as a robust and efficient control scheme capable of sustaining high tracking precision and stability under varying inertia and disturbance conditions.

From an implementation perspective, the proposed FPPC-API-NET relies on algebraic nonlinear transformations and PI-based control structures, which are computationally lightweight and suitable for real-time embedded automotive controllers. Although the present study focuses on simulation-based validation, the control architecture is well-suited for practical deployment in steering systems with limited computational resources. Future work will focus on extending the controller to incorporate full vehicle dynamics, including roll motion and load transfer, followed by real-time experimental validation on an actual RSV platform with embedded hardware to assess performance under real-world uncertainties.

ACKNOWLEDGMENT

The authors would like to thank Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA) for providing financial support under the Postgraduate Research Grant No. PGRS200348 and for laboratory facilities.

REFERENCES

[1] El Hajjami, L., Mellouli, E. M., Žuraulis, V., & Berrada, M. (2023). A novel robust adaptive neuro-sliding mode steering controller for autonomous ground vehicles. Robotics and Autonomous Systems, 170, 104557, doi:10.1016/j.robot.2023.104557.

[2] Zhou, X., Wang, Z., Shen, H., & Wang, J. (2022). Robust Adaptive Path-Tracking Control of Autonomous Ground Vehicles With Considerations of Steering System Backlash. IEEE Transactions on Intelligent Vehicles, 7(2), 315-325, doi:10.1109/TIV.2022.3146085.

[3] Chen, G., Yao, J., Hu, H., Gao, Z., & Hua, M. (2024). Hierarchical Coupling Motion Control of Distributed Vehicle Considering Vehicle Posture. IEEE Transactions on Vehicular Technology, 73(8), 10899-10915, doi:10.1109/TVT.2024.3369923.

[4] Liang, J., Feng, J., Lu, Y., Yin, G., Zhuang, W., & Mao, X. (2024). A Direct Yaw Moment Control Framework Through Robust T-S Fuzzy Approach Considering Vehicle Stability Margin. IEEE/ASME Transactions on Mechatronics, 29(1), 166-178, doi:10.1109/TMECH.2023.3274689.

[5] Dong, E., Wang, F., Li, B., & Zhang, L. (2022) Vehicle lateral stability control based on sliding model in high-speed sharp turning condition. In 2022 International Seminar on Computer Science and Engineering Technology (SCSET), 8-9 Jan. 2022 2022 (pp. 183-186). doi:10.1109/SCSET55041.2022.00051.

[6] Yim, S. (2020). Comparison among Active Front, Front Independent, 4-Wheel and 4-Wheel Independent Steering Systems for Vehicle Stability Control. Electronics, 9(5), 798, doi:10.3390/electronics9050798.

[7] Rsetam, K., Khawwaf, J., Zheng, Y., Cao, Z., & Man, Z. (2024). Fast Finite-Time Composite Controller for Vehicle Steer-by-Wire Systems with Communication Delays. World Electric Vehicle Journal, 15(4). doi:10.3390/wevj15040132

[8] Zhou, S., Li, Y., & Tong, S. (2024). Prescribed performance adaptive fuzzy output feedback control for steer-by-wire vehicle system with intermittent actuator faults. Neural Comput. Appl., 36(26), 16057–16070, doi:10.1007/s00521-024-09797-6.

[9] Zhang, J., & Li, J. (2019). Integrated vehicle chassis control for active front steering and direct yaw moment control based on hierarchical structure. Transactions of the Institute of Measurement and Control, 41(9), 2428-2440, doi:10.1177/0142331218801131.

[10] Liang, J., Yin, G., & Li, G. (2018) Robust H∞ Output-feedback Vehicle Yaw Control Using an Active Front Wheel Steering. In 2018 37th Chinese Control Conference (CCC), 25-27 July 2018 (pp. 7760-7764). doi:10.23919/ChiCC.2018.8483289.

[11] Alturky, L., & Khawwaf, J. (2025) Precision Tracking for Steer-by-Wire Systems Based on Fast Integral Terminal Sliding Mode Control. In 2025 5th International Conference on Emerging Smart Technologies and Applications (eSmarTA), 5-6 Aug. 2025 2025 (pp. 1-8). doi:10.1109/eSmarTA66764.2025.11132223.

[12] Yang, R., & Liao, D. (2025). Fault-Tolerant Control of Steer-by-Wire System Based on Sliding Mode Observer With Neural Network. IEEE Access, 13, 134831-134845, doi:10.1109/ACCESS.2025.3593492.

[13] Gao, Z., Liu, W., Wei, Z., & Guo, G. (2025). Adaptive Finite-Time Prescribed Performance Control of Vehicular Platoons With Multilevel Threshold and Asymptotic Convergence. IEEE Transactions on Intelligent Transportation Systems, 26(5), 6503-6513, doi:10.1109/TITS.2025.3536002.

[14] Nguyen, N. T., Le, D. T., Dang, V. T., Pham, V. H., Nguyen, D. H., Hoang, D. C., et al. (2022) Prescribed Tracking Performace for Lateral Control of an Autonomous Vehicle with High-Gain Observer. In 2022 11th International Conference on Control, Automation and Information Sciences (ICCAIS), 21-24 Nov. 2022 2022 (pp. 158-163). doi:10.1109/ICCAIS56082.2022.9990307.

[15] Wang, P., Feng, T., Song, C., Li, J., & Yang, S. X. (2025). A Study of the Stability of an Industrial Robot Servo System: PID Control Based on a Hybrid Sparrow Optimization Algorithm. Actuators, 14(2), 49, doi:10.3390/act14020049.

[16] Afrianto, A., Heryadi, B., Dewi, D. K., Wiranata, T., Rahman, H. A., Nurprasetio, I. P., et al. (2025). Experimental Investigation of PID Control Strategy for Three-Phase Induction Motor Speed Control System Implemented on a Scaled Railway Vehicle Roller Test Rig Driver. International Journal of Sustainable Transportation Technology, doi:10.31427/IJSTT.2025.7.1.1.

[17] Azahar, M. I. P., Irawan, A., & Ramli, M. S. (2020) Finite-Time Prescribed Performance Control for Dynamic Positioning of Pneumatic Servo System. In 2020 IEEE 8th Conference on Systems, Process and Control (ICSPC), 11-12 Dec. 2020 2020 (pp. 1-6). doi:10.1109/ICSPC50992.2020.9305755.

[18] Bian, M., Chen, L., Luo, Y., & Li, K. (2014). A Dynamic Model for Tire/Road Friction Estimation under Combined Longitudinal/Lateral Slip Situation (Vol. 1, SAE Technical Papers).

[19] Bhoraskar, A., & Sakthivel, P. (2017) A review and a comparison of Dugoff and modified Dugoff formula with Magic formula. In 2017 International Conference on Nascent Technologies in Engineering (ICNTE), 27-28 Jan. 2017 2017 (pp. 1-4). doi:10.1109/ICNTE.2017.7947898.

[20] Fawzan Khaizuran, F., & Irawan, A. (2020). Steering Angle Control of Rack Steering Vehicle using Antiwindup-PI-Control. International Journal of Electrical Engineering and Applied Sciences (IJEEAS), 3(1).

https://doi.org/10.31436/iiumej.v27i2.3991

Quantifying Activity of the Erector Spinae Longissimus Subgroup Using EMG During Prolonged Driving

MUHAMMAD ALFATIH ELTAYEB, NUR LIYANA AZMI*, AHMAD JAZLAN

Mechatronics Department, International Islamic University Malaysia, Kuala Lumpur, Malaysia

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

(Received: 2 October 2025; Accepted: 11 February 2026; Published online: 10 May 2026)

ABSTRACT: Prolonged driving has been associated with localized muscle fatigue, particularly in the lumbar region, posing safety and ergonomic concerns. This study investigates the dynamics of multivariate surface electromyography (sEMG) features to detect fatigue during extended driving sessions. EMG signals were recorded from the erector spinae muscles of seven subjects across two backrest inclination angles (100°, 110°) during real-world, multi-hour driving tasks. Spectrogram-based features, including mean frequency (MNF), median frequency (MDF), and root-mean-square (RMS), were extracted using a sliding-window method for analysis. Co-temporal trends in these features were analyzed to identify fatigue regions, marked by simultaneous decreases in MNF and MDF and/or increases in RMS. Paired t-tests were used to compare the two backrest inclination angles across subjects for each fatigue duration and fatigue event count. Results demonstrate that seat inclination angle significantly affects fatigue duration and number of events, with the 100° angle leading to notably more fatigue than 110° ($p<0.01$). This work could contribute a framework for time-resolved, EMG fatigue modeling in realistic driving scenarios.

ABSTRAK: Pemanduan terlalu lama dikaitkan dengan keletihan otot setempat, khususnya di bahagian lumbar, menimbulkan kebimbangan dari segi keselamatan dan ergonomik. Kajian ini mengkaji dinamika ciri permukaan multivarian elektromiografi (sEMG) bagi mengesan keletihan semasa sesi pemanduan jauh. Isyarat EMG direkodkan daripada otot erektor spina dari 7 subjek pada dua sudut kecondongan penyandar (100°, 110°) semasa aktiviti pemanduan sebenar berlangsung selama beberapa jam. Ciri berasaskan spektrogram, termasuk frekuensi purata (MNF), frekuensi median (MDF) dan purata punca kuasa dua (RMS) telah diekstrak menggunakan tetingkap-gelongsor. Corak serentak dalam ciri-ciri ini dianalisa bagi mengenal pasti kawasan keletihan, dibuktikan dengan penurunan MNF dan MDF serta/atau peningkatan RMS. Ujian-t berkembar digunakan bagi membandingkan dua sudut kecondongan penyandar merentas subjek bagi setiap tempoh keletihan dan bilangan kejadian. Dapatan kajian menunjukkan bahawa sudut kecondongan penyandar memberi kesan signifikan terhadap tempoh keletihan dan bilangan kejadian, dengan sudut 100° menghasilkan lebih banyak keletihan berbanding 110° ($p<0.01$). Kajian ini berpotensi menyumbang kepada satu rangka kerja pada model keletihan EMG beresolusi masa dalam senario pemanduan realistik.

KEY WORDS: *Prolonged Driving, Muscle Fatigue, Low Back Pain, EMG.*

1. INTRODUCTION

One critical factor affecting driving-induced low back pain is the driver's seat posture and seat inclination angle [1], as these directly affect muscle load and activation. The erector spinae longissimus subgroup plays a vital role in the stabilization of the trunk in this scenario. Muscle fatigue is defined as an exercise-induced decrease in the ability to produce force [1]. Prolonged

sitting while driving often leads to low back pain and muscle fatigue [2]. Some studies investigated both physical and mental fatigue while driving. Physiological muscular fatigue has been best assessed through spectral analysis of the muscle's electromyography (EMG) signal. A number of studies show that it results in a downward trend in both mean frequency (MNF) and median frequency (MDF) [3, 4].

In [5], actual driving was conducted to study the effect of seat design. The study found that the suspended seat delayed the onset of significant discomfort compared to soft and firm seats. Pre- and post- endurance static tests were used to quantify fatigue. The study focused on seat types and did not consider inclination angle. In [6], surface electromyography (sEMG) of the neck and shoulder of professional and non-professional drivers was analyzed. Significant changes in electrical activity were found to exist in the left deltoid, bilateral trapezius, and splenius capitis muscle groups of all subjects. This change, however, occurred between the 1st and 15th minutes of the short drive, and it was on a driving simulator.

The study in [7] quantified muscle fatigue using spectral analysis of sEMG during driving and achieved an accuracy of up to 85% with a random forest model. The study considered only one seat configuration, did not conduct an inter-subject variability analysis, and investigated the trapezius muscle. In [8], sEMG features were derived using the Wavelet Packet Transform (WPT) and the Continuous Wavelet Transform (CWT) to classify segments into three levels: comfortable, fatigue, and painful. The study was done on a simulation. The study in [8] analyzed the effects of neck balancing and lumbar support, but the driving scenario and duration were only 1 hour, and the analysis used RMS only. In [9], they analyzed the influence of different seat adjustments on driving fatigue using the AnyBody system. The study found that a backrest inclination of 10 to 20 degrees reduced muscle activation and lowered L4-L5 compressive forces during steering tasks, while also moderating loads during pedal operation.

Despite existing research using sEMG, there are limitations, such as driving duration, lack of seat angle variation, or reliance on simulations. Previously, in [10], we studied the overall averaged trends in the changes in MNF and MDF over a three-hour drive. However, some MNF and MDF trend results contradicted each other. In this study, the relationship among features of the EMG signals and the resulting fatigue indicators is further studied in a time-resolved manner: how each feature indicates fatigue, and when they simultaneously indicate fatigue.

This paper is organized as follows: the next part describes the methodology, experimental protocol, data acquisition, and processing pipeline for EMG feature extraction and fatigue detection. This is followed by the presentation of results, highlighting fatigue duration and event counts across the two seat-inclination angles. The discussion section then interprets these findings.

2. METHODOLOGY

2.1. Subjects

Seven healthy male subjects (mean age: 27.7 ± 5.3 years; mean weight: 64.3 ± 13.5 kg; mean height: 171.2 ± 4.0 cm) volunteered for the study. A sample size of 6 to 16 is common in similar studies [6,9,11], and this sample of seven subjects aligns with established research practice. Each subject completed three driving trials, each lasting approximately 3 hours, for a total of 63 hours. This study was approved by the IIUM Research Ethics Committee (IREC 2023-039). The subjects were: 1) aged between 20 and 30 years; and 2) free from symptoms

of low back and cervical pain prior to participating in the experiment. All subjects held a valid driver's license and provided written informed consent prior to their participation.

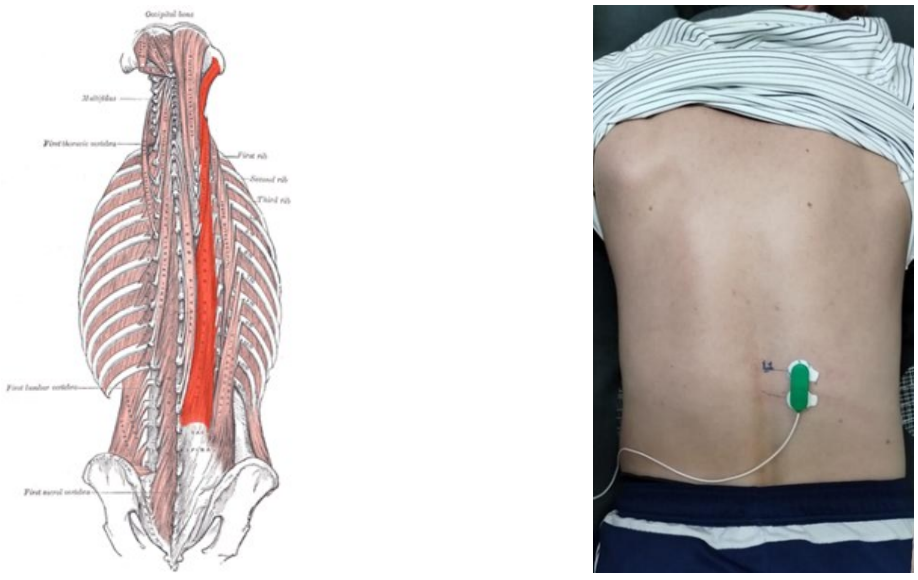
2.2. Experimental Setup

Before the experiment started, the seat backrest angle was set. The angle was measured between the pan and backrest as shown in Figure 1, using an inertial measurement unit (IMU) sensor (MPU6050), indicating the angle (θ) via an Arduino board. Each subject drove for three hours with θ = 100 and 110 degrees. The route consisted of a 73 KM lap through the section (Sungai Buloh - Ampang Jaya) of the Middle Ring Road 2, MRR2, Kuala Lumpur. Each subject was requested to drive at a maximum speed of 80km/h.



(a) Seat angle, θ (b) Seat angle set at 100° (c) Seat angle set at 110°

Figure 1. Seat angle, θ: the angle measured between the seat pan and the backrest.



(a) (b)

Figure 2. (a) The erector spinae group with the longissimus subgroup highlighted. (b) Sensor electrode placement.

To measure EMG, a Bitalino EMG sensor with an assembled BITalino Core BT/BLE board was used. Pre-gelled 24 mm polymer Ag/AgCl electrodes with a 20 mm inter-electrode distance were positioned vertically over the longissimus subgroup of the erector spinae muscle group. SENIAM guidelines dictated electrode placement at two finger-widths lateral to the spinous process of the L1 vertebra [11].

https://doi.org/10.31436/iiumej.v27i2.3991

The longissimus muscle subgroup is between the iliocostalis and spinalis. It is the largest of the three columns. It originates in the lumbar region and attaches to the lower 9 or 10 ribs between the tubercles and angles, the transverse processes of C2 – T12, and the mastoid process of the skull. It extends the vertebral column, as shown in Figure 2, and the overview of the experiment is illustrated in Figure 3. The EMG data were read by a computer through Bitalino's OpenSignals software, which connects to the board via Bluetooth. The sampling frequency used was 1kHz. The data analysis was carried out on MATLAB R2024b.

2.3. Signal Processing

2.3.1. Signal Preprocessing

Prior to analysis, the EMG signals were filtered using a bandpass filter with cutoff frequencies of 30 Hz and 450 Hz, as recommended in [12], to remove low-frequency artifacts and high-frequency noise. To mitigate power line interference (PLI) commonly observed at 50 Hz and its harmonics in Malaysia, notch filters were applied at 50 Hz and its harmonics (100, 150, 200, 250, 300, and 350 Hz) [13]. The DC offsets of the signals are then removed by subtracting their means.

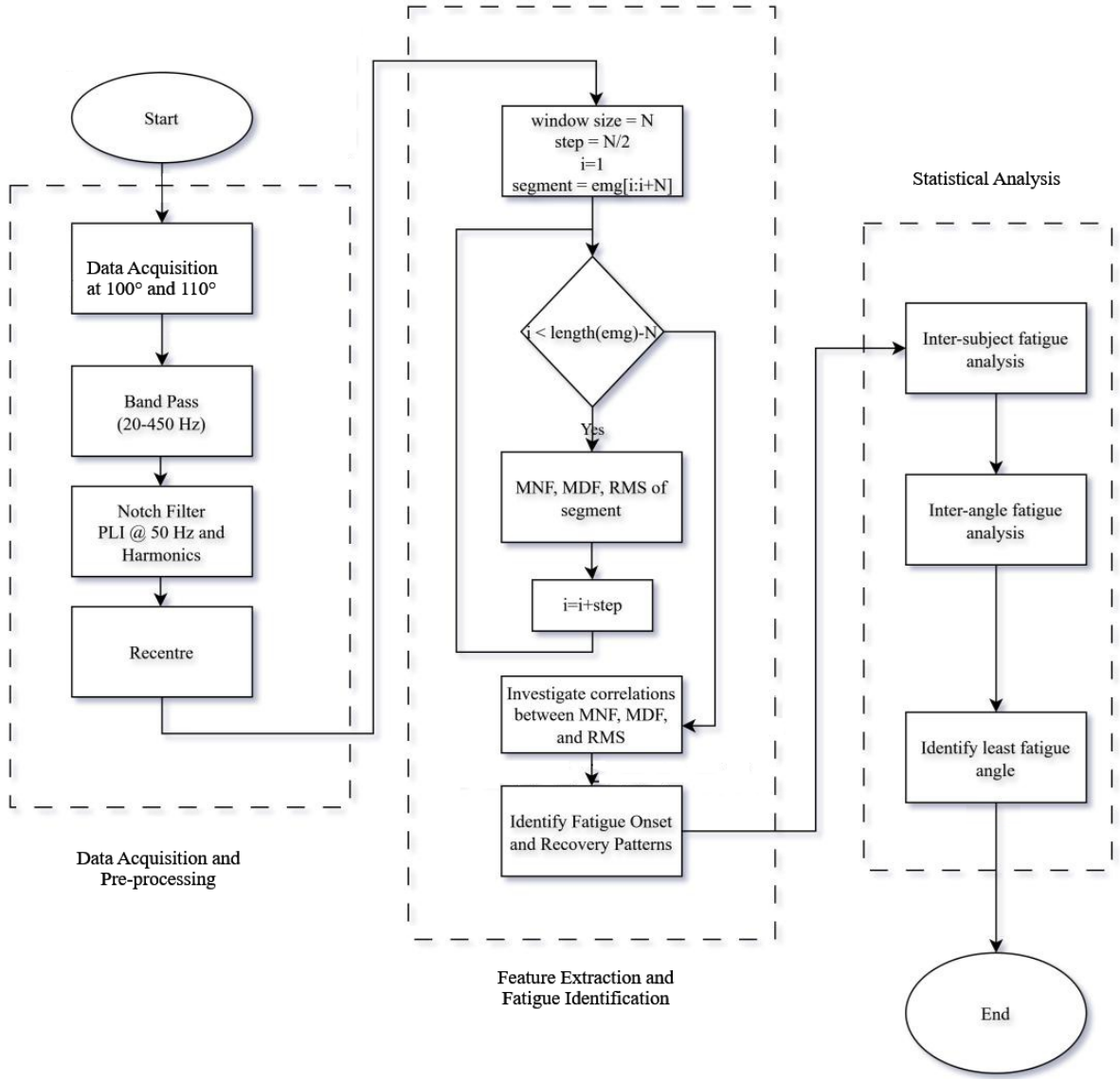


Figure 3. Flowchart of the fatigue assessment methodology.

https://doi.org/10.31436/iiumej.v27i2.3991

The signals were analyzed using a 1-second sliding window. A 1-second window was selected based on the recommendation in [14], which suggests a window size of approximately 500ms for active muscle contractions and up to 1s for low-intensity muscle activity, such as with the longissimus in the driving scenario. For each 1-second window, the root-mean-square (RMS) was calculated to quantify muscle activation. Windows with RMS values below the overall mean RMS were excluded from further analysis, as they typically represent periods of minimal muscle activity. Removing these low-activation segments reduces the impact of baseline noise and improves the signal-to-noise ratio, thereby increasing the accuracy of feature extraction.

2.3.2. Feature Extraction

From each window, the five extracted features were RMS, MNF, and MDF, which measure central tendencies. These were computed as follows:

$$RMS = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

$$MNF = \sum_{j=1}^{M} f_j P_j \,/\, \sum_{j=1}^{M} P_j \tag{2}$$

MDF is f_k such that

$$\sum_{j=1}^{k-1} f_j P_j < \frac{1}{2}\sum_{j=1}^{M} f_j P_j \text{ and } \sum_{j=1}^{k} P_j \geq \frac{1}{2}\sum_{j=1}^{M} P_j \tag{3}$$

x_i is the signal values within a specified moving window of size n. P_j and f_j are the power and the frequency at bin j, and M is the number of frequency bins.

2.4. Fatigue Markers

The features were first aggregated in 10-second averaging windows, with a sliding step of 5 seconds. Regions were marked as displaying fatigue if at least two of the following criteria were met:

1. A consistent downward trend in Mean Frequency (MNF)
2. A consistent downward trend in Median Frequency (MDF)
3. An increase or plateau in Root Mean Square (RMS)

For a minimum duration of 30 seconds (at least 6 consecutive windows) with at least 2 of the 3 conditions above met. The 30-second duration was selected to avoid noise effects at the low muscle activation levels used in [15,16].

For each trial, total fatigue duration was calculated by tallying the lengths of all identified fatigue episodes, and fatigue event count was defined as the number of distinct episodes detected. These summary metrics were then statistically analyzed.

2.5. Statistical Analysis

The total fatigue duration and number of fatigue events were summarized for each subject and seat angle. For each metric, the normality of paired differences between the angles (100°, 110°) was evaluated using the Lilliefors test. Paired t-tests will be used to compare the two angles when the normality assumptions are met; otherwise, the Wilcoxon signed-rank test will be used. Effect sizes were also calculated using Cohen's *d* to quantify the magnitude of differences and their practical significance.

3. RESULTS AND DISCUSSION

Patterns of regions indicating fatigue were intermittent. Figure 4 presents the results from a representative subject. No distinct trigger point is identified; instead, there is a gradual increase in the frequency of occurrence and duration of such regions. To provide a clearer view of short-term fluctuations, Figure 5 presents a zoomed-in segment of the same data, corresponding to the dotted box in Figure 4.

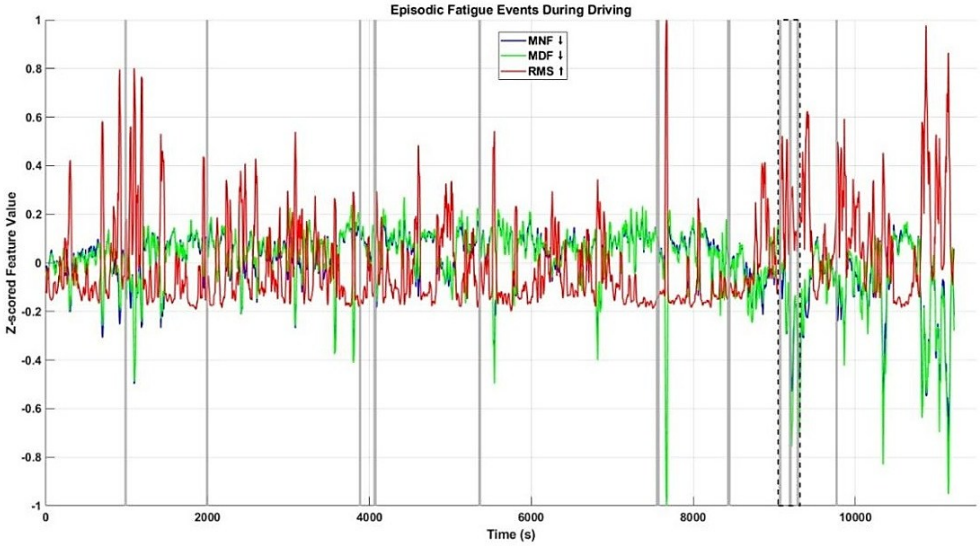


Figure 4. Intermittent regions where MDF, MNF and/or RMS indicate fatigue.

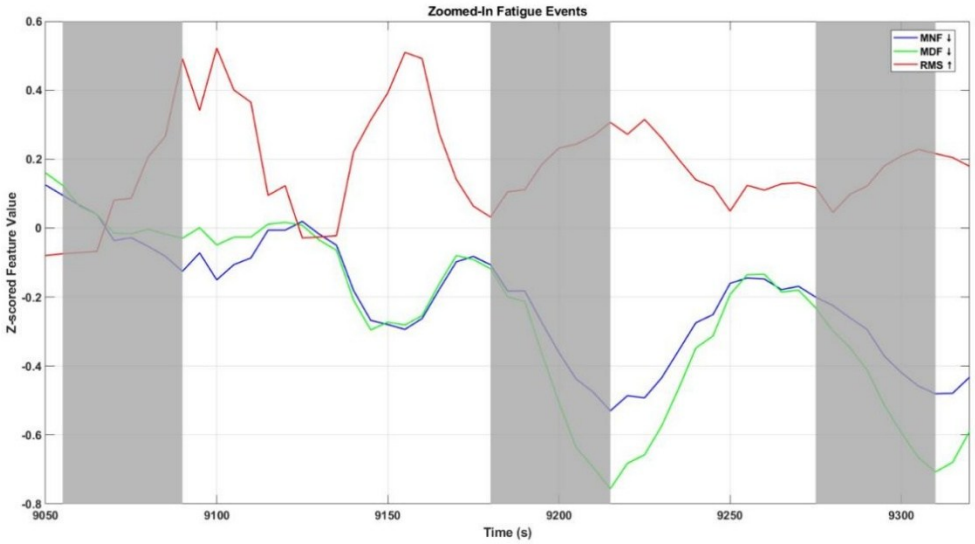


Figure 5. Zoomed-in view of Figure 4 ($9050 \leq t \leq 9320$ s) showing localized fatigue dynamics.

Table 1 presents the total fatigue duration and fatigue counts across all seven subjects and seat inclination angles. The mean fatigue duration for the two seat angles is shown in Figure 6: 1020.71 s at 100° and 661.43 s at 110 °.

https://doi.org/10.31436/iiumej.v27i2.3991

Table 1. Summary of total fatigue duration and fatigue counts across subjects and seat inclination angles

Subject	Seat Angle (°)	Fatigue Duration (s)	Fatigue Events (counts)
1	100	785	24
	110	395	11
2	100	1125	29
	110	835	22
3	100	610	17
	110	350	10
4	100	1480	42
	110	1165	33
5	100	890	26
	110	525	16
6	100	625	18
	110	265	8
7	100	1630	41
	110	1095	30

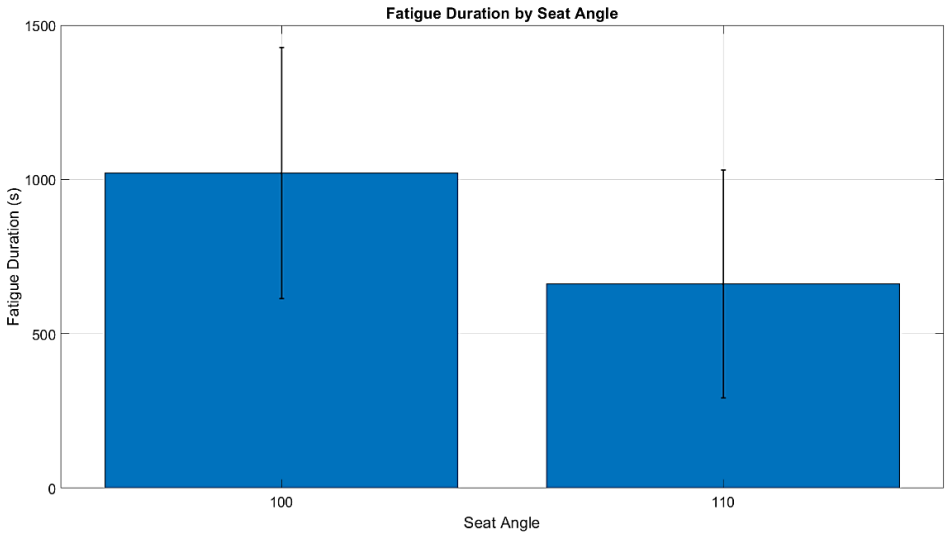


Figure 6. Total duration of fatigue regions for each seat angle.

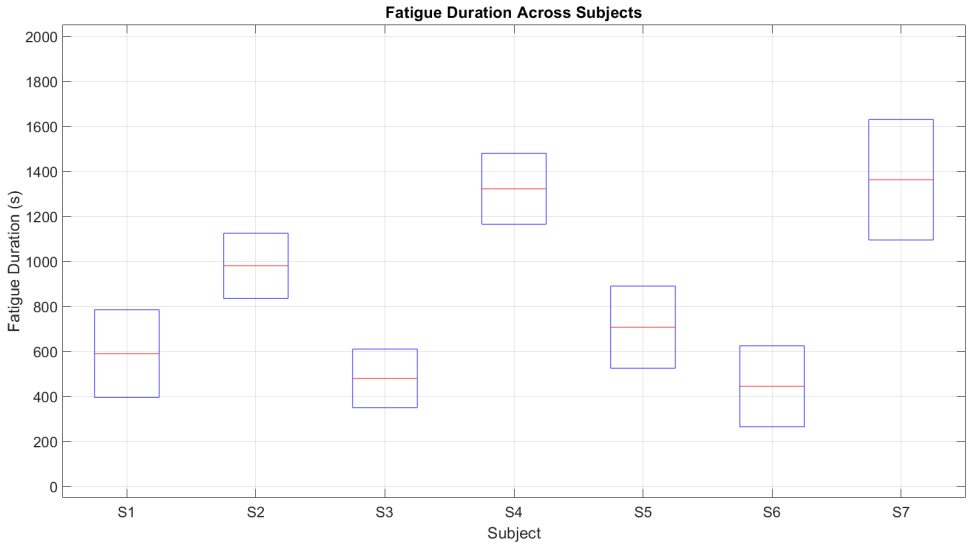


Figure 7. Total fatigue duration for each subject.

https://doi.org/10.31436/iiumej.v27i2.3991

Inter-subject comparisons showed significant differences in both fatigue duration and the number of fatigue events, with p-values of 0.039 and 0.046, respectively. Figure 7 shows fatigue duration across all seven subjects. Some subjects (e.g., Subjects 4 and 7) showed longer and more frequent fatigue periods than others. This makes them clearly distinct when compared pairwise with selected subjects with shorter durations (e.g., Subject 6). However, other pairs of profiles have overlapping distributions. However, this is not generalized among all pairs of subjects.

The normality of the paired differences between seat inclination angles (100° vs. 110°) was confirmed using the Lilliefors test (Duration: $p = 0.359$; Events: $p = 0.500$).

Results of the t-test showed that fatigue duration was significantly shorter at 110° compared to 100° ($t(6) = 10.576$, $p < 0.0001$, 95% CI). The effect size was extremely large (Cohen's d = 3.997), indicating a practical, robust reduction in fatigue duration at the more reclined posture.

Similarly, the number of fatigue events was significantly lower at 110° than at 100° ($t(6) = 11.783$, $p < 0.0001$, 95% CI), with a large effect size (Cohen's d = 4.453). The total duration of fatigue regions, subject-wise, is shown in Figure 8.

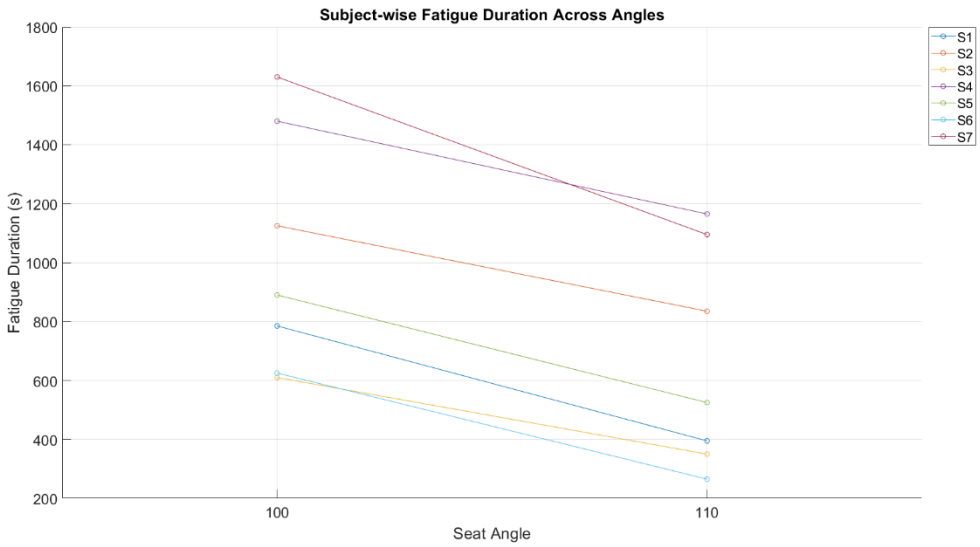


Figure 8. Subject-wise total fatigue regions duration vs. seat inclination angle.

This study found that the longissimus subgroup experienced less fatigue during prolonged driving at a seat inclination angle of 110° than at 100°. This is likely due to the effect of a more reclined backrest on reducing the load on the lumbar erector spinae and on load distribution, which in turn slows the spectral downshift (MNF or MDF) and limits the response in RMS increases [6, 17].

This study has some limitations. Since the analysis was limited to a single subgroup, the findings cannot be generalized to all lumbar muscles. Extending the investigation to include other muscle groups or subgroups may provide a more comprehensive understanding of their behavior under similar conditions. In addition, High-Density EMG (HDEMG) could offer deeper insights into spatial variations and simultaneous changes along the muscle itself. Finally, examining the synchronization of left and right subgroups may help clarify the role of the central nervous system (CNS) coordination in driving-related muscle fatigue.

4. CONCLUSION

This study found that a seat inclination angle of 110° was significantly more favorable for the erector spinae during prolonged driving compared to 100°. This finding may be applied in ergonomics to enhance driver comfort, for example, by optimizing lumbar support angles. Furthermore, this study contributes a framework for time-resolved, EMG fatigue modeling in realistic driving scenarios. In future studies, the use of HDEMG could provide more information on the muscle itself; recording background noise would improve noise cancellation, and examining the left-right muscle synergy could clarify how these muscles cooperate.

ACKNOWLEDGEMENT

The authors would like to express their gratitude to all participants in the experiment.

REFERENCES

[1] Wan JJ, Qin Z, Wang PY, Sun Y, Liu X (2017). Muscle fatigue: General understanding and treatment. Experimental & Molecular Medicine, 49(10):e384. https://doi.org/10.1038/emm.2017.194

[2] Chen C, Xiao B, He X, Wu J, Li W, Yan M (2024). Prevalence of low back pain in professional drivers: A meta-analysis. Public Health, 231:23–30. https://doi.org/10.1016/j.puhe.2024.03.007

[3] Roldán Jiménez C, Bennett P, Ortiz García A, Cuesta Vargas AI (2019). Fatigue detection during sit-to-stand test based on surface electromyography and acceleration: A case study. Sensors, 19(19):4202. https://doi.org/10.3390/s19194202

[4] Mader L, Herzberg M, Anders C (2024). Reliability of sEMG data of back muscles during static submaximal loading situations: Values and pitfalls. Journal of Electromyography and Kinesiology, 79:102947. https://doi.org/10.1016/j.jelekin.2024.102947

[5] Phinyomark A, Thongpanja S, Hu H, Phukpattaranont P, Limsakul C (2012). The usefulness of mean and median frequencies in electromyography analysis. Computational intelligence in electromyography analysis: a perspective on current applications and future challenges, pp 195. https://doi.org/10.5772/50639.

[6] Lecocq T, Chevrel C, Gorce P, Madeleine P (2022). Perceived discomfort and neuromuscular fatigue during long-duration real driving with different car seats. Frontiers in Sports and Active Living, 4:990748. https://doi.org/10.3389/fspor.2022.990748

[7] Azmi NL, Ghafar NAA, Nor KAM, Nordin NHD (2022). Classification of muscle fatigue during prolonged driving. ELEKTRIKA: Journal of Electrical Engineering, 21(3):40–46. https://doi.org/10.11113/elektrika.v21n3.376

[8] Tao X, Cheng B, Wang B, Zhang F, Li G, Chen C (2013). SEMG based recognition for lumbar muscle fatigue during prolonged driving. In Proceedings of the FISITA 2012 World Automotive Congress: Vehicle Design and Testing (I), 7:773–783.

[9] Menotti F, Labanca L, Laudani L, Giombini A, Pigozzi F, Macaluso A (2015). Activation of neck and low-back muscles is reduced with the use of a neck balance system together with a lumbar support in urban drivers. PLoS ONE, 10(10):e0141031. https://doi.org/10.1371/journal.pone.0141031

[10] Eltayeb MA, Azmi NL, Md Nor KA, Toha SF, Hashim A, Nordin NHD (2025). sEMG analysis of muscle fatigue during prolonged driving. PERINTIS EJournal, 15(1):15–21. Retrieved from https://perintis.org.my/ejournalperintis/index.php/PeJ/article/view/179

[11] Katsis CD, Ntouvas NE, Bafas CG, Fotiadis DI. (2004) Assessment of muscle fatigue during driving using surface EMG. *Proceedings of the 2nd IASTED International Conference on Biomedical Engineering*, 259–262.

[12] Hermens HJ, Freriks B, Merletti R, Stegeman D, Blok J, Rau G, Hägg G (1999). European recommendations for surface electromyography: Results of the SENIAM project. Roessingh Research and Development. Retrieved from http://seniam.org/

[13] De Luca CJ, Gilmore LD, Kuznetsov M, Roy SH (2010). Filtering the surface EMG signal: Movement artifact and baseline noise contamination. Journal of Biomechanics, 43(8):1573–1579. https://doi.org/10.1016/j.jbiomech.2010.01.027

[14] Ng CL, Reaz MBI, Ali SHBM, Crespo ML, Cicuttin A, Chowdhury MEH, Kiranyaz S, Kamal NB (2023). Powerline interference suppression of a textile-insulated capacitive biomedical sensor using digital filters. Measurement, 207:112425. https://doi.org/10.1016/j.measurement.2022.112425

[15] Butler HL, Hubley-Kozey CL, Kozey JW (2009). Electromyographic assessment of trunk muscle activation amplitudes during a simulated lifting task using pattern recognition techniques. Journal of Electromyography and Kinesiology, 19(6):e505–e512. https://doi.org/10.1016/j.jelekin.2008.09.010

[16] Conwit RA, Stashuk D, Suzuki H, Lynch N, Schrager M, Metter EJ (2000). Fatigue effects on motor unit activity during submaximal contractions. Archives of Physical Medicine and Rehabilitation, 81(9):1211–1216. https://doi.org/10.1053/apmr.2000.6975

[17] Song J, Choi YS, Lee S, Park D, Park J (2025). Changes in muscle oxygenation and activity during cumulative isometric muscle contraction: New insight into muscle fatigue. Frontiers in Physiology, 16:1559893. https://doi.org/10.3389/fphys.2025.1559893

[18] Majid NAA, Abdullah MFE, Jamaludin MS, Notomi M, Rasmussen J (2013). Musculoskeletal analysis of driving fatigue: The influence of seat adjustments. Advanced Engineering Forum, 10:373–378. https://doi.org/10.4028/www.scientific.net/aef.10.373

https://doi.org/10.31436/iiumej.v27i2.4021

Integration of Neutrosophic Methods into Adaptive Control of Nonlinear Systems Using Neuro-Fuzzy Networks With B-Spline Functions

OKSANA PORUBAY[1*], DILNOZA UMURZAKOVA[2]

[1]Department of Software Engineering and Cybersecurity, Faculty of Information Technology and Telecommunications, Fergana State Technical University, Fergana, Uzbekistan
[2]Department of Computer Engineering and Artificial Intelligence, Faculty of Information Technology and Telecommunications, Fergana State Technical University, Fergana, Uzbekistan

**Corresponding author: oksanaporubay@gmail.com*

(Received: 3 Nov 2025; Accepted: 19 January 2026; Published online: 10 May 2026)

ABSTRACT: This paper addresses the problem of adaptive control of nonlinear dynamic systems operating under parametric uncertainty, external disturbances, and partial or contradictory information about the system state – conditions under which classical linear Model Reference Adaptive Control (MRAC) and conventional neuro-fuzzy controllers exhibit degraded performance, slow adaptation, and oscillatory behavior. To overcome these limitations, a novel Neuro-Neutrosophic Model Reference Adaptive Controller (NN-MRAC) is proposed, implemented using a neutrosophic neuro-fuzzy network with B-spline basis functions. The key innovation of the proposed approach lies in integrating neutrosophic logic into the adaptive control architecture by explicitly using a three-component uncertainty representation – truth, indeterminacy, and falsity – which enables robust control synthesis in the presence of incomplete, noisy, and conflicting data. In contrast to traditional neuro-fuzzy controllers, the proposed NN-MRAC combines localized B-spline approximation with neutrosophic weighting of local models and an adaptive decomposition into lower-dimensional submodels, effectively mitigating the curse of dimensionality and reducing computational complexity. Comparative simulation studies with a classical linear MRAC demonstrate that the proposed controller reduces the mean-square tracking error by approximately 59%, decreases overshoot by more than 3 times, shortens the transient response time by nearly 1.8 times, and lowers control energy consumption by about 18%. The results confirm that the proposed neuro-neutrosophic MRAC ensures stable, smooth, and energy-efficient control in the presence of noise and deep uncertainty, making it a promising solution for intelligent control of complex nonlinear systems.

ABSTRAK: Kajian ini mencadangkan masalah kawalan adaptif bagi sistem dinamik bukan linear yang beroperasi pada ketidakpastian parameter, gangguan luaran, serta maklumat separa atau bercanggah mengenai keadaan sistem, di mana kaedah linear klasik Model Kawalan Suai Rujukan (MRAC) dan kawalan konvensional neural-fuzi menunjukkan prestasi terhad, kadar penyesuaian perlahan, dan tingkah laku berayun. Bagi mengatasi kekangan ini, satu Model Kawalan Suai Rujukan Neuro-Neutrosofik (NN-MRAC) yang baharu dicadangkan, dilaksanakan berasaskan rangkaian neural-fuzi neutrosofik dengan fungsi asas alur-B. Inovasi utama pendekatan ini terletak pada pengintegrasian logik neutrosofik ke dalam seni bina kawalan suai melalui penggunaan perwakilan terkecuali pada tiga komponen ketidakpastian – kebenaran, ketidakpastian, dan kepalsuan – membolehkan sintesis kawalan yang teguh dalam keadaan data tidak lengkap, bising, dan bercanggah. Berbeza dengan pengawal neural-fuzi tradisional, NN-MRAC yang dicadangkan menggabungkan penghampiran setempat berasaskan alur-B dengan pemberat neutrosofik bagi model setempat serta penguraian suai kepada submodel berdimensi rendah, sekaligus mengurangkan

kerumitan pengiraan dan mengatasi masalah "kutukan dimensi". Kajian simulasi perbandingan dengan MRAC linear klasik menunjukkan bahawa cadangan kawalan mencapai pengurangan ralat min kuasa dua (MSE) kira-kira 59%, penurunan lebihan (overshoot) lebih daripada tiga kali ganda, pemendekan masa tindak balas sementara hampir 1.8 kali, serta pengurangan penggunaan tenaga kawalan sekitar 18%. Keputusan ini mengesahkan bahawa MRAC neuro-neutrosofik yang dicadangkan mampu memastikan kawalan stabil, licin, dan cekap tenaga di bawah keadaan hingar dan ketidakpastian mendalam, menjadikannya satu penyelesaian berpotensi pada sistem kawalan pintar bukan linear yang kompleks.

KEYWORDS: *Adaptive Control, Neutrosophic Logic, MRAC, Neuro-Fuzzy Network, B-Spline.*

1. INTRODUCTION

Modern dynamic systems operate under conditions of high uncertainty, nonlinearity, and variability of external factors, which significantly complicates the synthesis of efficient control systems. Traditional methods of linear adaptive control, including classical models such as MRAC (Model Reference Adaptive Control) [1, 2], demonstrate good performance only within a narrow range of operating conditions where the object can be approximated by a linear model. However, under the influence of unpredictable disturbances, stochastic fluctuations, and parametric deviations, their effectiveness sharply decreases. In such circumstances, it becomes necessary to develop intelligent systems capable of self-learning, self-organization, and real-time adaptation. The significance of this problem is particularly pronounced in application domains such as electric power systems, robotics, autonomous and cyber-physical systems, where nonlinear dynamics, parameter drift, sensor noise, and incomplete state information are intrinsic characteristics of the operating environment. In such systems, insufficient robustness of adaptive controllers may lead not only to performance degradation but also to instability, increased energy consumption, and reduced operational safety. Therefore, the development of adaptive control strategies capable of explicitly handling uncertainty and partial state awareness represents an important and practically relevant scientific challenge.

Among various areas of intelligent control, neuro-fuzzy systems hold a special place, as they combine the heuristic principles of fuzzy logic with the universal approximating capabilities of neural networks. These hybrid structures allow modeling nonlinear dependencies without requiring an exact mathematical description of the object. An important element in such systems is the selection of basis functions, which determine the smoothness and stability of the approximation. In this context, B-spline functions possess several advantages, including localized approximation, low computational complexity, and the ability to adaptively adjust the function shape to data during the learning process. Recent studies have demonstrated that intelligent adaptive controllers based on ANFIS, neuro-PID structures, and deep neuro-fuzzy architectures can significantly improve control accuracy for nonlinear systems compared to classical linear MRAC. These approaches exploit data-driven learning and nonlinear approximation to compensate for modeling errors and disturbances. However, most existing methods implicitly assume that the available data are reliable and internally consistent, and uncertainty is typically modeled only through noise or bounded parameter variations. Despite their effectiveness, conventional neuro-fuzzy and neural adaptive controllers exhibit several fundamental limitations. First, classical fuzzy systems operate with a single degree of membership, which does not explicitly represent indeterminacy or contradictory information. Second, neural-network-based controllers often behave as black-box models, limiting interpretability and increasing the risk of unstable adaptation under sparse or conflicting data. Third, an increase in the number of input variables typically leads to a rapid

growth in computational complexity, known as the curse of dimensionality, which restricts real-time implementation in embedded and resource-constrained systems. These limitations indicate a clear research gap in adaptive control under complex and poorly structured uncertainty conditions.

Nevertheless, even conventional neuro-fuzzy models encounter significant challenges when operating under deep uncertainty, where parts of the data are contradictory, and other parts are incomplete or unreliable. To overcome these limitations, neutrosophic logic, an extension of classical fuzzy logic proposed by F. Smarandache, has been actively developed in recent years. In this approach, each variable is described by three independent components: truth (T), indeterminacy (I), and falsity (F) [3-5]. This triad enables the accounting for ambiguous and conflicting situations characteristic of complex technological systems and ensures more accurate decision-making under incomplete information. The integration of neutrosophic principles into the MRAC controller based on B-spline neuro-fuzzy networks forms a new generation of neutrosophic neuro-fuzzy adaptive control systems characterized by high robustness to uncertainty and the ability to learn in a context-dependent manner. This architecture not only expands the system's adaptive capabilities but also enables intelligent analysis of control actions that account for fuzzy, probabilistic, and neutrosophic information.

The scientific novelty of the proposed approach lies in the development of a hybrid neuro-neutrosophic MRAC that uses B-spline functions to construct local models with varying degrees of certainty (T, I, F), thereby enabling adaptive adjustment of the network's weight coefficients during system operation. The study also proposes a submodel decomposition method that reduces computational complexity and eliminates the "curse of dimensionality." Thus, the developed concept represents a step toward creating a new generation of intelligent, adaptive controllers that combine the precision of mathematical models, the flexibility of fuzzy systems, and the robustness of neutrosophic logic. The practical application of this approach is particularly relevant for control tasks in power systems, robotic systems, autonomous complexes, and other nonlinear dynamic systems operating under uncertainty.

In contrast to existing adaptive control approaches, the proposed method introduces an explicit neutrosophic representation of uncertainty within the adaptive control loop, enabling simultaneous consideration of truth, indeterminacy, and falsity in control decision-making. The use of B-spline basis functions provides a localized and smooth approximation of nonlinear dynamics, while the proposed submodel decomposition strategy ensures scalability and real-time feasibility. As a result, the developed neuro-neutrosophic MRAC differs from conventional MRAC, ANFIS, and deep neuro-fuzzy controllers by offering improved robustness to deep uncertainty, reduced computational complexity, and enhanced interpretability of control actions.

2. RELATED WORKS

Adaptive control of nonlinear systems has been extensively studied over the past decades, with classical Model Reference Adaptive Control (MRAC) forming the foundation of many modern approaches [1, 2]. While MRAC provides guaranteed stability under certain assumptions, its performance deteriorates significantly in the presence of strong nonlinearities, unmodeled dynamics, and uncertainty in system parameters. To address these limitations, intelligent control methods based on artificial neural networks and fuzzy logic have been proposed.

Adaptive Neuro-Fuzzy Inference Systems (ANFIS) and neuro-PID controllers have demonstrated improved approximation capability and adaptability compared to linear MRAC,

particularly for nonlinear plants with unknown dynamics [8, 9]. These methods exploit data-driven learning to compensate for modeling inaccuracies and external disturbances. More recently, deep neuro-fuzzy and hybrid intelligent controllers have been introduced to further enhance control performance, robustness, and convergence speed by leveraging multi-layer learning architectures and nonlinear feature extraction. Despite these advances, most existing intelligent adaptive controllers rely on classical fuzzy logic or probabilistic uncertainty modeling, in which uncertainty is represented implicitly through noise, bounded disturbances, or parameter variations. Such representations do not explicitly distinguish between reliable information, incomplete knowledge, and contradictory data. As a result, these controllers may exhibit unstable adaptation, reduced interpretability, or degraded performance when operating under deep uncertainty and partial awareness of system states.

Neutrosophic logic, introduced as a generalization of fuzzy logic, provides a more expressive framework for uncertainty modeling by representing each variable using three independent components: truth, indeterminacy, and falsity [3-5]. Several recent studies have explored neutrosophic sets and inference mechanisms in decision-making, optimization, and control-related applications [14, 15]. However, the integration of neutrosophic logic into adaptive control systems, particularly in MRAC architectures and neuro-fuzzy networks, remains limited and underexplored.

In contrast to existing approaches, the present study integrates neutrosophic logic directly into the adaptive control loop of an MRAC framework using B-spline-based neuro-fuzzy networks. Unlike conventional neuro-fuzzy and deep learning controllers, the proposed method explicitly accounts for epistemic uncertainty and contradictory information while maintaining computational efficiency and interpretability through local B-spline approximation and submodel decomposition. This positions the proposed neuro-neutrosophic MRAC as a distinct and complementary contribution to the current state of the art in intelligent adaptive control of nonlinear systems.

3. THE PROBLEM OF ADAPTIVE CONTROL OF NONLINEAR SYSTEMS

Modern nonlinear systems are characterized by high uncertainty, parameter variability, and stochastic disturbances and structural nonlinearity [6]. In such systems, traditional adaptive control methods based on classical MRAC algorithms lose their effectiveness for several reasons. First, they assume that the structure of the controlled object is known in advance and can be described by an accurate mathematical model. Second, they rely on the assumption of smooth, linear dynamics in the vicinity of the operating point, which does not hold in most real installations. Finally, traditional methods do not account for the influence of multidimensional interactions and cannot effectively handle incomplete or contradictory data [7].

The typical problems that arise when applying linear MRAC systems to complex dynamic objects can be identified as follows:

1) Structural uncertainty: the parameters of the object model change over time, and its dynamics cannot be approximated by a single linear model.
2) Nonlinear dependencies between inputs and outputs: the interaction of multiple factors (temperature, load, response time, etc.) leads to nonlinear effects that cannot be compensated for by linear control.
3) Data uncertainty: measurements are often accompanied by noise, incompleteness, or even logical inconsistency, especially when working with distributed systems.

4) Computational complexity problem: an increase in the number of input variables leads to an exponential growth in model dimensionality (the "curse of dimensionality" effect), making the learning process unstable and slow.

Attempts to extend classical MRAC methods using Artificial Neural Networks (ANNs) and Fuzzy Logic have led to the emergence of hybrid neuro-fuzzy controllers [8, 9]. However, these methods also retain several limitations: first-type fuzzy systems are sensitive to noise and do not account for uncertainty within the membership functions themselves; neural networks, in turn, require large volumes of training data and often suffer from overfitting. As a result, there is a need for a more universal approach that can combine the advantages of both methods while maintaining interpretability and robustness to uncertainty.

At this point, the neutrosophic theory of control, grounded in neutrosophic logic, comes to the forefront [10, 11]. Unlike classical fuzzy logic, where the degree of membership is defined by a single value within the range [0,1], the neutrosophic approach considers a triad of values: $T(x)$, $I(x)$, $F(x)$, where T is the degree of truth, I is the degree of indeterminacy, and F is the degree of falsity.

This allows the description of states in which information is contradictory or partially unreliable, which is critically important for complex technical systems. Thus, the use of a neuro-neutrosophic MRAC controller implementing adaptation based on B-spline functions makes it possible to simultaneously:

- account for the nonlinearity of dynamics and parameter uncertainty;
- localize approximation and thereby reduce computational complexity;
- describe not only the degree of membership but also the level of uncertainty in the data;
- maintain model interpretability through an IF-THEN rule structure;
- ensure smooth and continuous control actions during transient processes.

The main problem addressed in this study is the development of a stable and computationally efficient approach to controlling nonlinear systems under model and input uncertainties, when traditional algorithms lose their ability to adapt adequately [12]. To overcome these limitations, a hybrid neuro-neutrosophic MRAC architecture is proposed that combines the principles of neural networks, fuzzy logic, and neutrosophic analysis.

Based on the above analysis, this study formally addresses the following problem: developing an adaptive control approach for nonlinear dynamic systems that ensures stable, accurate tracking performance under deep uncertainty, including parametric variability, external disturbances, and incomplete or contradictory information about the system state, while maintaining computational efficiency and interpretability. The limitations of existing MRAC, neuro-fuzzy, and neural adaptive controllers indicate the necessity of a control architecture that can simultaneously approximate nonlinear dynamics, explicitly represent different types of uncertainty, and adapt in real time without excessive computational burden. Therefore, the objective of this study is to design a hybrid neuro-neutrosophic Model Reference Adaptive Controller that integrates B-spline-based local approximation with neutrosophic uncertainty representation.

The main contributions of this work are summarized as follows:

1) formulation of a neuro-neutrosophic MRAC framework that explicitly incorporates truth, indeterminacy, and falsity into the adaptive control loop;

2) development of a B-spline-based neuro-fuzzy architecture enabling localized approximation and smooth control actions;
3) introduction of a submodel decomposition strategy to mitigate the curse of dimensionality and ensure real-time feasibility;
4) demonstration of improved robustness, stability, and energy efficiency through comparative simulation studies.

4. SOLUTION METHOD

4.1. Overview of the Method

Consider a discrete nonlinear dynamic system represented as an input-output model, which is described by the following expression:

$$z(t+1) = b_0 z(t) + b_1 z(t-1) + \cdots + b_{n_z-1} z(t-n_z+1) + \Psi[v(t), v(t-1), \dots, v(t-r+1)] + \eta(t) \quad (1)$$

where $v(t)$ is the control variable, $z(t)$ is the system output variable, n_z and r are the known model orders for the output and input, respectively, $\eta(t)$ is a random noise characterized by independent, identically distributed values with zero mean and variance σ^2, $\Psi[\cdot]$ is a continuously differentiable nonlinear function that admits a Taylor series expansion in the vicinity of the operating point.

In this case, the following conditions are met:

$$\partial\Psi/(\partial v(t+1)) = 0; \partial\Psi/\partial v(t) \neq 0. \quad (2)$$

From expression (2), it follows that the Jacobian of the function $\Psi[\cdot]$ exists and is continuous. Then, equation (1) can be represented in operator form as:

$$P(q^{-1})z(t) = q^{-1}\Psi[v(t), v(t-1), \dots, v(t-r+1)] + \eta(t), \quad (3)$$

where q^{-1} is the inverse time shift operator.

The local linearized model of system (1) in the vicinity of the operating point v(t) can be expressed as follows:

$$\bar{P}(q^{-1})z(t) = q^{-1}\bar{Q}(q^{-1})v(t) + \eta(t), \quad (4)$$

where $\bar{P}(q^{-1})$ and $\bar{Q}(q^{-1})$ are polynomials in the operator q^{-1} of the corresponding orders n_z' and n_v'. The coefficients of these polynomials depend on the current state of the system and are determined in the vicinity of the operating point S(t) [13].

For ease of analysis, we introduce a state vector containing the observed output values of the system:

$$\begin{cases} x_1(t) = z(t-n_z+1), \\ x_2(t) = z(t-n_z+2), \\ \vdots \\ x_{n_z}(t) = z(t). \end{cases} \quad (5)$$

Taking into account the introduced notations, the original expression (1) can be transformed into the form of a state space:

$$\begin{cases} X(t+1) = AX(t) + B\Psi[v(t), v(t-1), \dots, v(t-r+1)], \\ z(t) = CX(t), \end{cases} \quad (6)$$

where $X = [x_1, x_2, \dots, x_{n_z}]^T \in \mathbb{R}^{n_z}$ is the state vector, and the matrices A, B, and C have the following form:

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \cdots \\ 0 & 0 & \dots & 0 & 1 \\ b_0 & b_1 & \dots & b_{n_z-2} & b_{n_z-1} \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \quad C = [0 \ 0 \ \cdots \ \cdots \ 1].$$

This representation enables moving from the "input–output" form to a state-space model, which is convenient for stability analysis and adaptive controller design [13]. However, under real-world conditions, the parameters b_i and the function $\Psi[\cdot]$ are not defined precisely, but rather as interval-fuzzy and neutrosophic quantities that include not only the degree of truth (T) but also the degrees of indeterminacy (I) and falsity (F).

Taking into account neutrosophic logic, model (6) can be extended to the following form:

$$X(t+1) = A(T,I,F)X(t) + B(T,I,F)\Psi[v(t), v(t-1), \dots, v(t-r+1)], \quad (7)$$

where each matrix includes parameters expressed in neutrosophic form $A(T,I,F) = A_T + A_I i + A_F f$ which makes it possible to describe uncertain, partially contradictory, and probabilistic relationships among the system states.

This generalization provides the foundation for constructing a neuro-neutrosophic adaptive control architecture that combines the advantages of fuzzy networks, neural network approximation, and neutrosophic interpretation of uncertainty.

4.2. Neutrosophic Interpretation and Architecture of the Neuro-Fuzzy Network

The conceptual architecture of the proposed neuro-neutrosophic fuzzy network utilizing B-spline basis functions is shown in Fig. 1.

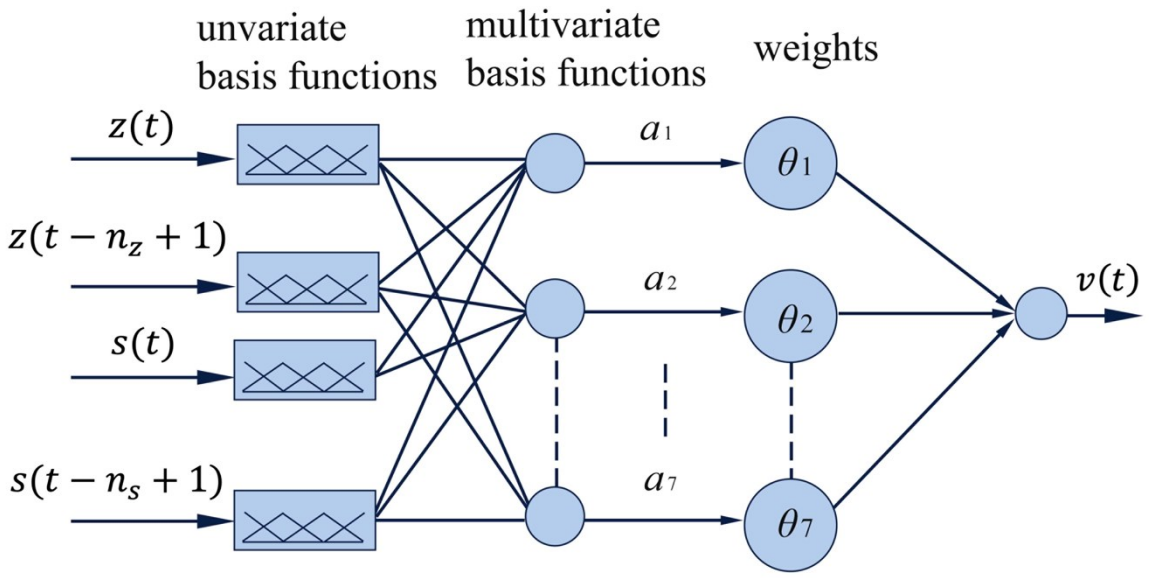


Figure 1. Structure of the neutrosophic B-spline neuro-fuzzy network.

A vector of signals is fed to the network input

$$[z(t), z(t-1), \dots, z(t-n_z+1), s(t-n_s+1)], \quad (8)$$

where $z(t)$ are the current and previous values of the output variable, and $s(t)$ is the reference (or control) signal.

Each of these inputs undergoes neutrosophic fuzzification, which includes not only the standard membership degree $\mu(x)$ but also the parameters of indeterminacy $I(x)$ and falsity $F(x)$, forming the neutrosophic membership triad $\langle T(x), I(x), F(x)\rangle$ [14, 15].

Fuzzification is performed using one-dimensional B-spline basis functions, which provide smooth and adaptive partitioning of the input data range into overlapping intervals. Unlike classical membership functions, each B-spline function in the neutrosophic version reflects the range of uncertainty through an additional coefficient I_i, which adjusts its shape and weight depending on the reliability of the training data.

The resulting membership function values form a neutrosophic feature vector, which is fed into the network's linear combination layer. At this stage, information aggregation is performed according to the rule:

$$y(t) = \sum_{i=1}^{q} W_i \cdot \Phi_i^{(p)}(\xi) \cdot T_i - F_i \cdot \delta_i, \tag{9}$$

where $\Phi_i^{(p)}(\xi)$ is a B-spline function of order p, T_i, I_i, F_i are neutrosophical reliability coefficients, W_i are network weights, δ_i is a correction term that reduces the influence of false knowledge.

The design process of the neutrosophic B-spline neuro-fuzzy network includes the determination of the following parameters:

- the order of the basis functions p, which defines the degree of smoothness of the approximation;
- the range of variation of input and output variables;
- the number of internal nodes q, i.e., the number of B-spline functions used for data fuzzification;
- the neutrosophic reliability coefficients $\langle T_i, I_i, F_i\rangle$, which regulate the level of confidence in local models.

The adjustment of the network weights $W = \{w_1, w_2, \dots, w_q\}$ is performed based on a training dataset using the neutrosophic gradient descent method, where the loss function accounts for data uncertainty:

$$E = \frac{1}{2} \sum_t (z(t) - \hat{z}(t))^2 \cdot (1 + I_t - F_t). \tag{10}$$

Thus, the higher the uncertainty of a training example, the smaller its contribution to the gradient correction, which prevents overfitting in the presence of noisy or contradictory data.

4.3. Properties and Advantages of Neutrosophic B-Spline Functions

The B-spline basis functions used in the neutrosophic architecture are defined by the recursive relation:

$$\Phi_i^{(p)}(\xi) = \frac{\xi - \lambda_{i-1}}{\lambda_{i+p-1} - \lambda_{i-1}} \Phi_{i-1}^{(p-1)}(\xi) + \frac{\lambda_{i+p} - \xi}{\lambda_{i+p} - \lambda_i} \Phi_i^{(p-1)}(\xi), \tag{11}$$

where $\Phi_i^{(0)}(\xi) = 1$, if $\lambda_{i-1} \le \xi < \lambda_i$, otherwise $\Phi_i^{(0)}(\xi) = 0$.

These functions possess the following neutrosophically interpreted properties:

1) Locality and contextual adaptivity. Each basis function $\Phi_i^{(p)}(\xi)$ is active only within a limited interval $[\lambda_{i-1}, \lambda_{i+p}]$, which ensures local information storage. In the neutrosophic modification, a reliability parameter T_i is introduced, which determines the strength of the response in each local region depending on the reliability of the data.
2) Normalization and output stability. For any input, the following condition holds:

$$\sum_{i=1}^{q} \Phi_i^{(p)}(\xi) T_i \equiv 1, \forall \xi \in [\xi_{min}, \xi_{max}], \quad (12)$$

which prevents network imbalance during training and ensures output stability in the presence of contradictory data.

3) Smoothness and learning consistency. The basis functions belong to the class $C^{(p-2)}$, ensuring continuity up to the $(p-2)$-th order. This guarantees smooth control actions even in the presence of noise or measurement errors.
4) Boundedness and semantic robustness of inference. Since the network output represents a neutrosophically weighted sum of local approximators, the resulting control is not only bounded but also reflects the model's "level of confidence."

Considering these properties, the integration of neutrosophic logic into the architecture of a B-spline neuro-fuzzy network enables the transition from traditional fuzzy modeling [16] to an intelligent representation of uncertainty, where each data component is described not merely by a degree of membership, but by three cognitive dimensions (T, I, F). This approach provides more robust and adaptive control under changing external conditions [17].

The proposed neutrosophic neuro-fuzzy network, illustrated in Fig. 1, is constructed based on a set of type-1 fuzzy rules that define the relationship between the system's input and output variables. Each rule R_i has the following form:

R_i: IF $z(t)$ is Negative Large, ..., and $z(t-n_z+1)$ is Negative Medium,

and $s(t)$ is Positive Medium, THEN $v(t)$ is Positive Medium.

Here, $z(t)$ denotes the system output, and $s(t)$ represents the reference signal. The set of such rules forms a hybrid reasoning system in which each local knowledge region reflects its own "neutrosophic context," that is, it includes the parameters of truth (T), indeterminacy (I), and falsity (F). This allows the extension of traditional fuzzy reasoning by introducing additional degrees of uncertainty, which is particularly important for controlling complex nonlinear systems with variable structures.

The output action of the network $v(t)$ is described by the aggregated expression:

$$v(t) = \Lambda^T(\xi)\, \Omega, \quad (13)$$

where $\xi(t) = [\, z(t), z(t-1), \dots, z(t-n_z+1), s(t), \dots, s(t-n_s+1)\,]^T$ is the input vector of observed variables, $\Omega = [\, \omega_1, \omega_2, \dots, \omega_P\,]^T$ is a vector of network weight coefficients, $\Lambda(\xi)$ is a vector of basis functions formed during the fuzzification process.

The number of network parameters P depends on the number of internal nodes w_i and the order of spline functions p_i for each input parameter:

$$P = \prod_{i=1}^{n_\xi} (w_i + p_i), \quad (14)$$

where $n_\xi = n_z + n_s$ – is the dimension of the input vector.

Each component function $\Lambda_i(\xi)$ is defined as a product of one-dimensional B-spline membership functions:

$$\Lambda_i(\xi) = \prod_{l=1}^{n_\xi} \phi_{A_l^i}(\xi_l), i = 1,2,\dots,P, \quad (15)$$

where $\phi_{A_l^i}(\xi_l)$ – is the membership function for the l-th coordinate of the input vector.

In terms of neutrosophic logic, each function $\varphi_{A_l^i}(\xi_l)$ can be further expressed as a triplet:

$$\phi_{A_l^i}(\xi_l) = \{T_l^i(\xi), I_l^i(\xi), F_l^i(\xi)\},$$

which makes it possible to account not only for the degree of membership but also for the levels of data indeterminacy and inconsistency within the fuzzy interpretation.

Then, the resulting control action can be written as:

$$v(t) = \sum_{i=1}^{P} \omega_i \Lambda_i(\xi) = \sum_{i=1}^{P} \omega_i \prod_{l=1}^{n_\xi} \phi_{A_l^i}(\xi_l). \quad (16)$$

Such a representation enables the system to be described by a set of local linear models, combined within a unified neutrosophic-fuzzy architecture. Each local model adapts to its own data subset, while the interaction among them realizes hybrid reasoning under uncertainty. Thus, the proposed approach integrates the classical MRAC structure with neutrosophic logic mechanisms, ensuring stable, interpretable behavior of the adaptive controller under nonlinear disturbances.

4.4. Training and Adaptation of the Neuro-Neutrosophic Fuzzy Network

The training process of the neuro-neutrosophic fuzzy network aims to adjust the weight coefficients and membership parameters that determine the importance of local reasoning rules [18, 19]. During adaptation, the target error function is minimized, which reflects the mismatch between the system's reference and actual outputs. The update scheme of the weight parameters $\Omega = [\omega_1, \omega_2, \dots, \omega_P]^T$ is defined by the recurrent expression:

$$\omega_i(t+1) = \omega_i(t) - \rho \frac{\partial \mathcal{J}}{\partial \omega_i} = \omega_i(t) - \rho\, \varepsilon(t) \frac{\partial \varepsilon(t)}{\partial \omega_i}, i = 1,2,\dots,P, \quad (17)$$

where ρ is the learning rate (adaptive step), $\mathcal{J} = \frac{1}{2}\varepsilon^2(t)$ is the loss function, $\varepsilon(t) = z_r(t) - z(t)$ is the tracking error between the reference and current outputs.

The derivative of the gradient with respect to the weight is determined considering the output sensitivity to the control action:

$$\frac{\partial \varepsilon(t)}{\partial \omega_i} = \frac{\partial z(t)}{\partial v(t)} \frac{\partial v(t)}{\partial \omega_i} = \frac{\partial z(t)}{\partial v(t)} \Lambda_i(\xi), \quad (18)$$

where $\Lambda_i(\xi)$ is the activated B-spline membership function at the current system state.

From the viewpoint of neutrosophic logic, each weight coefficient ω_i is interpreted not as a fixed parameter, but as a neutrosophic triad:

$$\omega_i = \{\omega_i^T, \omega_i^I, \omega_i^F\},$$

where ω_i^T represents the true component of knowledge (the reliable part of approximation), ω_i^I is the indeterminacy (incompleteness of information during training), ω_i^F is the falsity component approximation errors or noise).

https://doi.org/10.31436/iiumej.v27i2.4021

Such a representation makes it possible not only to minimize the error but also to quantitatively control the degree of uncertainty during the network adaptation process.

Due to the local nature of the B-spline bases, the input space of the network is divided into many non-overlapping regions of activity Q, each corresponding to a subset of active rules. The number of such regions is determined by the expression:

$$Q = (w_z + 1)^{n_z}(w_s + 1)^{n_s}, \quad (19)$$

where w_z and w_s are the numbers of internal approximation nodes for the variables z(t) and s(t), respectively, n_z and n_s are the orders of the corresponding components of the input vector.

At any given moment, only one subspace $Q_k \subset Q$is activated, which significantly reduces computational costs. For the active region, the number of non-zero elements in the fuzzified feature vector is defined as:

$$P' = p_z^{n_z} p_s^{n_s}, \quad (20)$$

where p_z and p_s – are the orders of the B-spline functions for the corresponding variables.

Thus, the training of the neuro-neutrosophic fuzzy network occurs simultaneously on three levels:

1) *Physical level*: adjustment of weights and approximation coefficients.
2) *Logical level*: refinement of fuzzy rules through updating of membership functions.
3) *Neutrosophic level*: correction of the degrees of truth, indeterminacy, and falsity, ensuring the stability of training even under noise and data incompleteness.

Such a multilevel approach makes it possible not only to achieve rapid training of the network in local regions of the input space but also to enable the self-organization of the structure of neuro-neutrosophic fuzzy rules, thereby significantly increasing the model's interpretability and robustness in adaptive control tasks for nonlinear systems.

4.5. Properties of the Multidimensional Neutrosophic-Fuzzy Regulator

The proposed neutrosophic-fuzzy adaptive controller represents a hybrid system consisting of two interconnected subsystems:

1) a static nonlinear part based on B-spline basis functions, which performs a topologically consistent mapping of the input variables;
2) an adaptive linear part, in which the parameters (weights) are adjusted as data are received, similar to the mechanisms of self-learning neural networks.

This approach allows the proposed controller to be interpreted as a self-organizing intelligent system that not only approximates the object's nonlinear dynamics but also evaluates the reliability of its responses through the neutrosophic component – the triad of truth, indeterminacy, and falsity (T, I, F).

Traditional fuzzy controllers (for example, of the Takagi–Sugeno type) are built as two-dimensional controllers with inputs the tracking error and its rate of change. Their logic is close to that of an experienced operator: if the error is large and the system deviates from the setpoint, then the control action should be increased to compensate.

Analytically, such controllers can be regarded as a combination of:

- a global multidimensional relay element, providing discrete switching between fuzzyfication regions;
- and a local nonlinear PID controller, implementing smooth control within each region.

In the case of extending the number of input variables to three or more, the structure remains the same, and the output control becomes equivalent to the sum of several local linear control submodels.

The proposed neutrosophic-fuzzy controller generalizes this principle, allowing each local model to be neutrosophically weighted, i.e., to account for the reliability of its parameters when computing the control signal.

Due to the nonlinearity of both the controlled object and the neutrosophic-fuzzy controller, achieving global stability of the closed-loop system is extremely difficult. However, under certain conditions, it is possible to ensure local system stability in the vicinity of the operating point.

Taking into account model (6) and the control action (16), the state space of the closed-loop system can be represented as:

$$\Xi(t+1) = A\Xi(t) + +B\Psi\left[\sum_{i=1}^{P}\omega_i\prod_{l=1}^{n_\xi}\phi_{A_l^i}(\xi_l(t)), \sum_{i=1}^{P}\omega_i^{(-1)}\prod_{l=1}^{n_\xi}\phi_{A_l^i}(\xi_l(t-1)), \dots, \sum_{i=1}^{P}\omega_i^{(-(r-1))}\prod_{l=1}^{n_\xi}\phi_{A_l^i}(\xi_l(t-r+1))\right], \quad (21)$$

where $\omega_i^{(-1)}, \dots, \omega_i^{(-(r-1))}$ are the weights obtained at previous training iterations, reflecting the temporal dependence of adaptation and the effect of the system's cognitive memory.

Let the reference point $s(t) = 0$, and $\bar{\Xi} = 0$ be the equilibrium point of the system. For the stability analysis, introduce a Lyapunov function defined on the compact S:

$$V\{\Xi(t)\} = \frac{1}{2}E^T E, \quad (22)$$

where $E = \Xi - \bar{\Xi} = \Xi$ – is the state error vector.

Then the Lyapunov function can be written as:

$$V\{\Xi(t)\} = \frac{1}{2}\Xi^T\Xi = \frac{1}{2}\{z^2(t-n_z+1) + \dots + z^2(t)\}, \quad (23)$$

and its value at the next step:

$$V\{\Xi(t+1)\} = \frac{1}{2}\{z^2(t-n_z+2) + \dots + z^2(t+1)\} \geq 0. \quad (24)$$

For sufficiently small changes in the weight coefficients $\Delta\omega_i$ such that $|\Delta\omega_i| \leq \delta, \forall i$, the change of the Lyapunov function can be written as:

$$\Delta V = V\{\Xi(t+1)\} - V\{\Xi(t)\} \approx \sum_i \frac{\partial V\{\Xi(t+1)\}}{\partial\omega_i}\Delta\omega_i = z(t+1)\sum_i \frac{\partial z(t+1)}{\partial v(t)}\frac{\partial v(t)}{\partial\omega_i}\Delta\omega_i. \quad (25)$$

Substituting the expressions for weight updates and the error gradient:

$$\omega_i(t+1) = \omega_i(t) - \gamma\frac{\partial J}{\partial\omega_i} = \omega_i(t) - \gamma\,\varepsilon(t)\,\frac{\partial\varepsilon}{\partial\omega_i}, \frac{\partial\varepsilon}{\partial\omega_i} = \frac{\partial z}{\partial v}\frac{\partial v}{\partial\omega_i} = \frac{\partial z}{\partial v}\Lambda_i(\xi), \quad (26)$$

We obtain:

$$\Delta V = -\gamma\, z^2(t+1)\sum_i\left(\frac{\partial z(t+1)}{\partial v(t)}\right)^2(\Lambda_i(\xi))^2 \leq 0. \quad (27)$$

Since $\Delta V \leq 0$, the system tends to equilibrium provided that the learning step γ satisfies $\gamma \leq \lambda_{\min}$, ensuring the smallness of the parameter increments $\Delta\omega_i$. Despite possible slow convergence, this guarantees local stability of the neutrosophic-fuzzy system in the vicinity of the stationary point.

In the context of neutrosophic logic, the process of adapting the weights $\omega_i$can be considered as the evolution of the triad (T, I, F), where: T is the degree of truth of the system response, I is the measure of indeterminacy (information entropy), F is the degree of falsity of the control decision.

Minimization of the Lyapunov functional $V\{\Xi(t)\}$ is equivalent to the reduction of indeterminacy I and the increase of truth T during training. Thus, the stability of the system in the classical Lyapunov sense has a cognitive interpretation in the neutrosophic space, namely, the convergence of reasoning toward truth through the minimization of indeterminacy in the solution space.

4.6. Algorithm Implementation and Simulation Setup

This subsection summarizes the practical implementation of the proposed neuro-neutrosophic MRAC and specifies the simulation conditions used for performance evaluation. The algorithmic workflow of the controller is organized as follows:

- At the initialization stage, the structure of the neuro-neutrosophic fuzzy network is defined, including the number and order of B-spline basis functions, the initial values of adaptive weights, and the neutrosophic parameters corresponding to truth (T), indeterminacy (I), and falsity (F). The reference model parameters and the initial conditions of the controlled system are also specified.
- At each sampling instant, the system output is measured and compared with the reference model output to compute the tracking error. Based on the current system state and the error signal, the input variables are processed by the B-spline-based fuzzification mechanism, yielding local activation functions and corresponding neutrosophic membership values. The control signal is generated by aggregating the weighted outputs of the local neuro-neutrosophic submodels, accounting for degrees of truth, indeterminacy, and falsity.
- The adaptive update of the network parameters is performed using a gradient-based learning law driven by the tracking error. This adaptation mechanism enables real-time adjustment of the controller parameters in response to changes in system dynamics, parametric uncertainty, and external disturbances. The algorithm iterates continuously until the predefined simulation horizon is reached.
- All simulations are carried out in MATLAB/Simulink using a fixed sampling period of Ts = 0.01 s. The nonlinear plant dynamics are subjected to parametric uncertainty by varying the nominal system parameters within ±20% of their baseline values. External disturbances are modeled as bounded stochastic noise. The reference model is chosen as a stable second-order linear system to ensure smooth, bounded tracking behavior. These simulation settings are chosen to reflect realistic real-time control conditions and to ensure reproducibility of the obtained results.

5. EXAMPLE OF SOLUTION

Let us consider a discrete nonlinear second-order system described by the equation:

$$z(t) = 0.3\, z(t-1) + 0.6\, z(t-2) + [v(t-1)]^{\frac{1}{3}} + \eta(t), \tag{28}$$

where $\eta(t)$ is a normally distributed white noise with zero mean and variance 0.1^2.

This system exhibits pronounced nonlinearity with respect to the control input and stochastic uncertainty, making it a suitable object for testing the efficiency of the neutrosophic-fuzzy MRAC controller.

5.1. Linear Reference MRAC

To begin, let us apply the classical MRAC, which uses the input vector

$$\xi(t) = [\, z(t), z(t-1), z(t-2), s(t)\,],$$

where $s(t)$ is the reference (setpoint) signal in the form of a rectangular wave.

Parameter adaptation is performed according to the gradient descent law:

$$\begin{aligned} \omega_1(t+1) &= \omega_1(t) - \gamma\, \varepsilon(t)\, s(t)\, [v(t-1)]^{-\frac{2}{3}}, \\ \omega_2(t+1) &= \omega_2(t) - \gamma\, \varepsilon(t)\, z(t)\, [v(t-1)]^{-\frac{2}{3}}, \\ \omega_3(t+1) &= \omega_3(t) - \gamma\, \varepsilon(t)\, z(t-1)\, [v(t-1)]^{-\frac{2}{3}}, \\ \omega_4(t+1) &= \omega_4(t) - \gamma\, \varepsilon(t)\, z(t-2)\, [v(t-1)]^{-\frac{2}{3}}. \end{aligned} \tag{29}$$

The initial parameters were chosen as $\omega_1^0 = \omega_2^0 = \omega_3^0 = 0.1$, with the learning rate $\gamma = 0.2$. The output signal of the linear MRAC exhibits high variability and significant oscillations, due to the linear model's inability to approximate the nonlinear relationship between the input and output (Fig. 2).

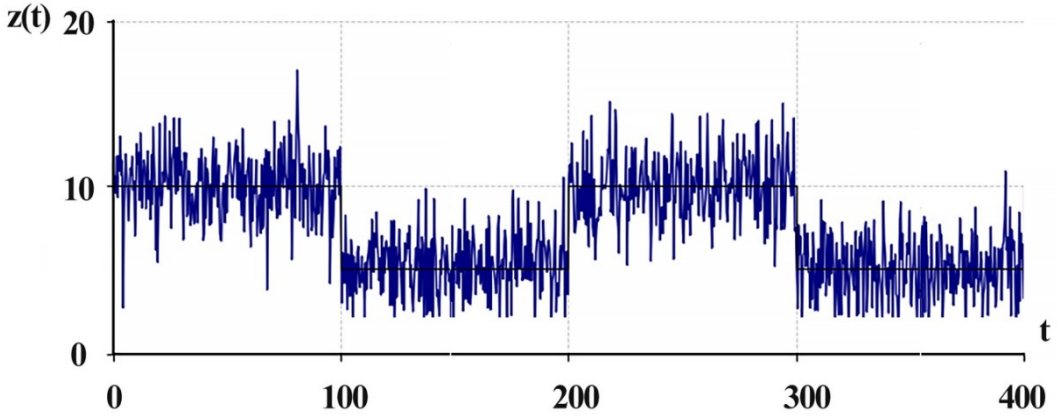


Figure 2. Closed-loop output of the nonlinear system under control of the linear MRAC.

5.2. Neutrosophic-Fuzzy MRAC

To eliminate the aforementioned drawbacks, an MRAC based on a neutrosophic-fuzzy network was implemented, using the same input vector structure $\xi(t)$ as the linear controller. Each input variable is fuzzified using two B-spline basis functions corresponding to the linguistic variables “low” and “high.” In this case:

$$v_z(t) = v_{z(t-1)} = v_{z(t-2)} = v_s(t) = 2, w_z(t) = w_{z(t-1)} = w_{z(t-2)} = w_s(t) = 0.$$

According to expression (14), the total number of network weights is:

$$P = 2^4 = 16.$$

The ranges of the input variables $z(t), z(t-1), z(t-2), s(t)$ are defined within the interval [0, 12]. The initial weights $\omega_i(0)$ are set the same as in the linear MRAC, and their updating is performed according to the rule:

$$\omega_i(t+1) = \omega_i(t) - \gamma\, \varepsilon(t)\, \Lambda_i(\xi(t))\, [v(t-1)]^{-\frac{2}{3}}, i = 1,2,\ldots,16. \quad (30)$$

Simulation results show that the neutrosophic-fuzzy MRAC yields significantly smaller oscillations and faster convergence than its linear counterpart (Fig. 3).

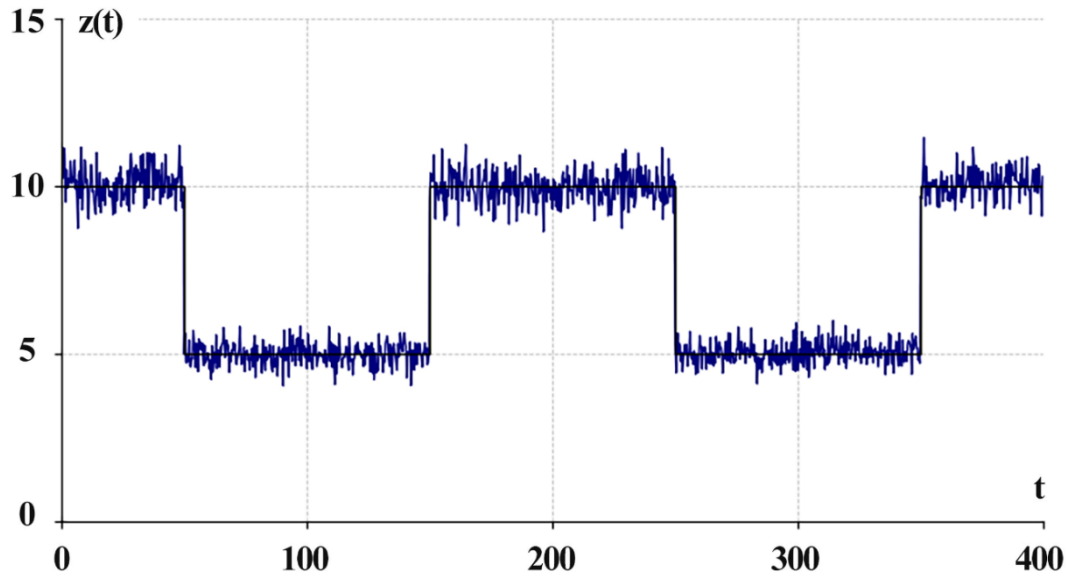


Figure 3. Closed-loop output of the system using the neutrosophic-fuzzy MRAC.

5.3. Improvement of Approximation and Submodule Decomposition

To further increase the accuracy of approximation, five B-spline functions were used for each variable, representing the linguistic states: “positive large,” “positive medium,” “zero,” “negative medium,” and “negative large.” A complete four-dimensional network would require 625 basis nodes, with 2 to the fourth power, or 16 local functions activated at each step. To reduce computational costs, the model was additively decomposed into submodules:

$$v(t) = \sigma_1[z(t), z(t-1)] + \sigma_2[z(t-2), s(t)]. \quad (31)$$

This structure allows the controller to be implemented as a combination of two two-dimensional neutrosophic-fuzzy subnetworks (Fig. 4), each using only 50 weights. This significantly reduces model complexity while maintaining high accuracy and stability.

https://doi.org/10.31436/iiumej.v27i2.4021

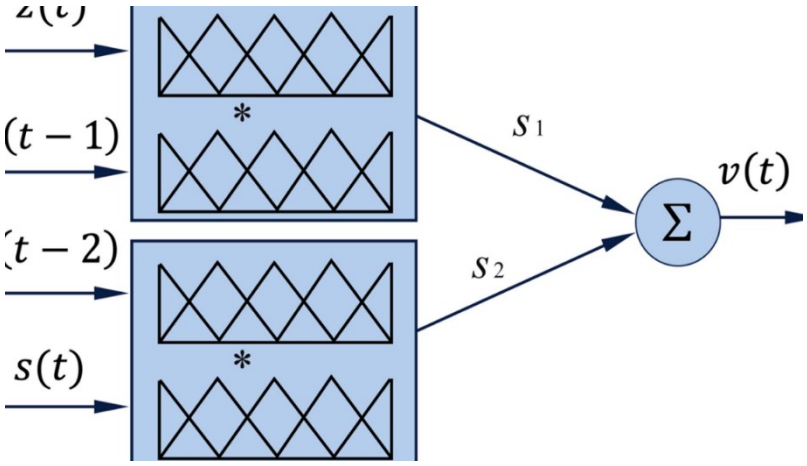


Figure 4. Neutrosophic-fuzzy controller implemented through smaller submodels.

5.4. Simulation Results

As the fuzzification resolution increases, the steady-state oscillations almost completely disappear (Fig. 5). For a quantitative evaluation of control performance, the integral of the squared control signal was calculated:

$$J = \sum_t v^2(t), \tag{32}$$

where $v(t)$ is the control signal.

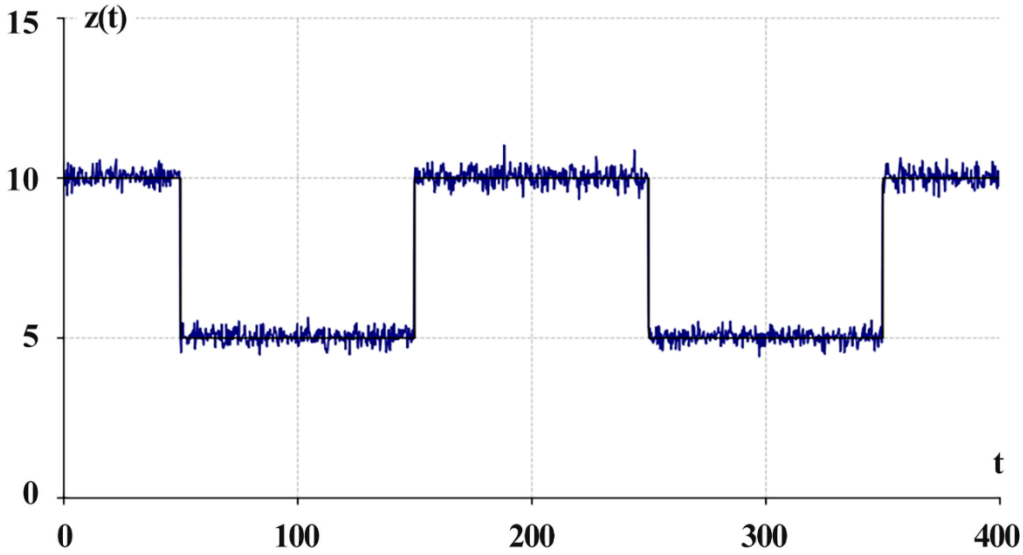


Figure 5. Closed-loop output under the submodular neutrosophic-fuzzy MRAC.

The neutrosophic-fuzzy MRAC based on B-splines demonstrated approximately an 18% reduction in control effort compared to the classical MRAC, indicating smoother actuator influence and improved energy efficiency.

Within the framework of neutrosophic analysis, the reduction of control signal energy and stabilization of the output response are interpreted as an increase in truth (T) and a decrease in indeterminacy (I) within the triad (T, I, F), where F represents the degree of falsity of decisions, which decreases during learning. Thus, the system cognitively "refines" its knowledge about the object, minimizing contradictions and reaching a stable state of reasoning.

6. NUMERICAL RESULTS AND COMPARATIVE ANALYSIS

6.1. Simulation Parameters

To verify the efficiency of the developed neutrosophic-fuzzy MRAC, a numerical experiment was conducted on model (28) with the following parameters:

- Simulation interval: $t \in [0,200]$;
- Discretization step: $\Delta t = 0.01$;
- Noise $\eta(t) \sim \mathcal{N}(0, 0.1^2)$;
- Learning rate: $\gamma = 0.2$;
- Number of basis functions per input: 5;
- Spline order: $p = 3$;
- Initial weights: $\omega_i(0) = 0.1$.

The comparison was carried out between the Linear MRAC, implemented according to Eq. (29), and the Neutrosophic-Fuzzy MRAC (NF-MRAC), based on the B-spline architecture Eq. (30).

6.2. Dynamics of Transient Processes

Figure 6 presents a comparison of the transient characteristics for both types of controllers under identical disturbance input and reference signal $s(t)$. The graphs show that the linear MRAC exhibits noticeable spikes and oscillations during the first 40 cycles, whereas the neutrosophic-fuzzy MRAC provides a smoother approximation with a reduced transient time and smaller overshoot.

The mean square error (MSE) was calculated using the expression:

$$E_{mse} = \frac{1}{N} \sum_{t=1}^{N} [z(t) - s(t)]^2, \tag{33}$$

where N is the length of the observation sample.

The results of the calculations are summarized in Table 1.

Table 1. Comparison of key control performance indicators

Type of Controller	Mean Squared Error (MSE)	Transient Tim t_p(cycles)	Overshoot (%)
Linear MRAC	0.0217	43	18.4
NF-MRAC (B-spline)	0.0089	24	5.7

From the table, it follows that the use of the B-spline neutrosophic-fuzzy structure reduced the tracking error by approximately 59%, decreased the overshoot by more than 3 times, and accelerated system settling by almost 1.8 times.

https://doi.org/10.31436/iiumej.v27i2.4021

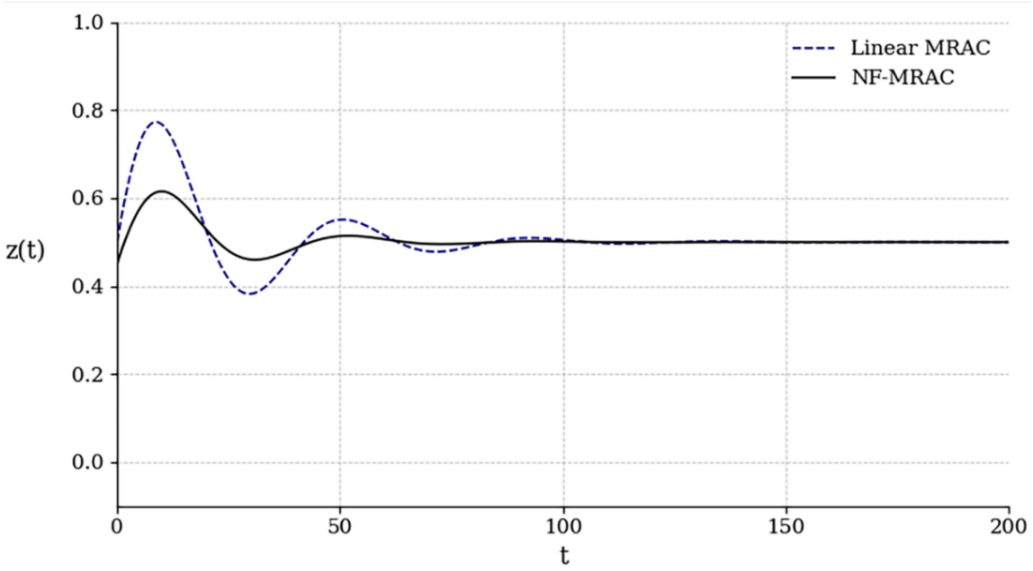


Figure 6. Comparison of transient processes of the linear and neutrosophic-fuzzy MRAC.

6.3. Analysis of the Control Action

To assess the controller's smoothness and energy efficiency, the control energy was calculated using the expression (32). The results showed that the NF-MRAC consumes approximately 18% less control energy than the classical MRAC, which is consistent with the smoother dynamics of the B-spline basis functions. Graphically, this is confirmed by the smaller amplitude of the control signal while maintaining a comparable response speed (Fig. 7).

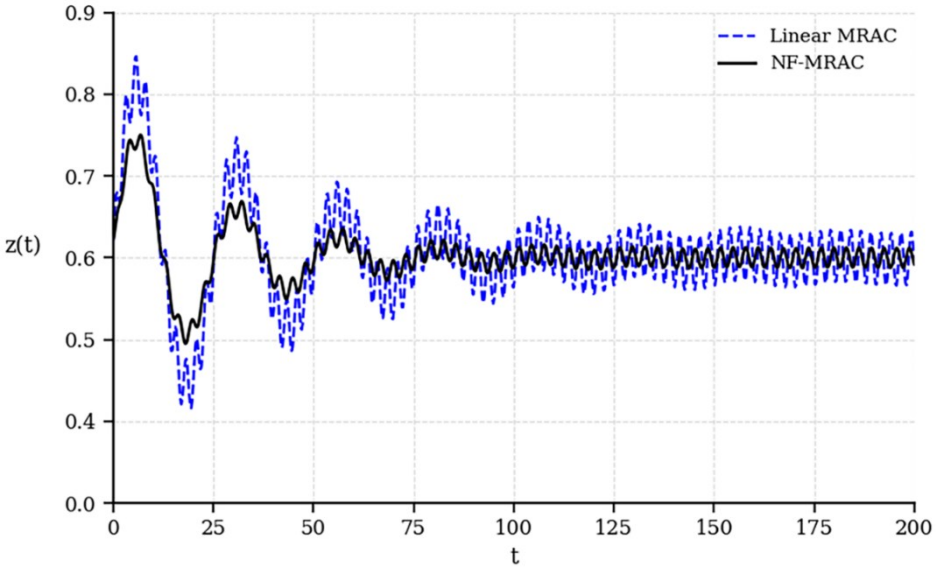


Figure 7. Comparison of control actions for MRAC and NF-MRAC.

6.4. Neutrosophic Analysis of Stability and Adaptation

From the standpoint of neutrosophic logic, the training process of the network can be viewed as a cognitive evolution of the triad (T, I, F), where: T – degree of truth, reflecting the correctness of the system dynamics approximation; I – indeterminacy, associated with data variability and stochastic noise; F – falsity, characterizing the model error and incorrect approximations. During iterative learning, the following dynamic transformation is carried out:

https://doi.org/10.31436/iiumej.v27i2.4021

$$(T, I, F)_{k+1} = (T_k + \Delta_T, I_k - \Delta_I, F_k - \Delta_F), \quad (34)$$

where $\Delta_T > 0, \Delta_I > 0, \Delta_F > 0$ are the increments that depend on the magnitude of the error $\varepsilon(t)$and the adaptation step γ.

The analysis shows that the neutrosophic-fuzzy MRAC possesses the property of cognitive stabilization – the system not only minimizes the error but also reduces internal indeterminacy, thereby bringing the control process closer to a state of maximum reliability ($T \to 1, I \to 0, F \to 0$).

6.5. Comparison with Modern Intelligent Controllers

For a comprehensive assessment of the efficiency of the proposed neutrosophic-fuzzy MRAC (NF-MRAC), a comparative analysis was conducted with three modern intelligent control systems:

- ANFIS (Adaptive Neuro-Fuzzy Inference System),
- Neuro-PID controller,
- Deep Fuzzy Controller (based on a convolutional neuro-fuzzy architecture).

The results are summarized in Table 2.

Table 2. Comparison of the Efficiency of Intelligent Controllers

Type of Controller	Mean Squared Error (MSE)	Transient Time t_p(cycles)	Overshoot (%)	Control Energy E_u	Noise Robustness
Linear MRAC	0.0217	43	18.4	1.00	Medium
Neuro-PID	0.0142	35	10.9	0.95	Good
ANFIS	0.0111	32	9.6	0.91	Good
Deep Fuzzy	0.0098	28	7.3	0.89	Very High
NF-MRAC (B-spline)	0.0089	24	5.7	0.82	Very High

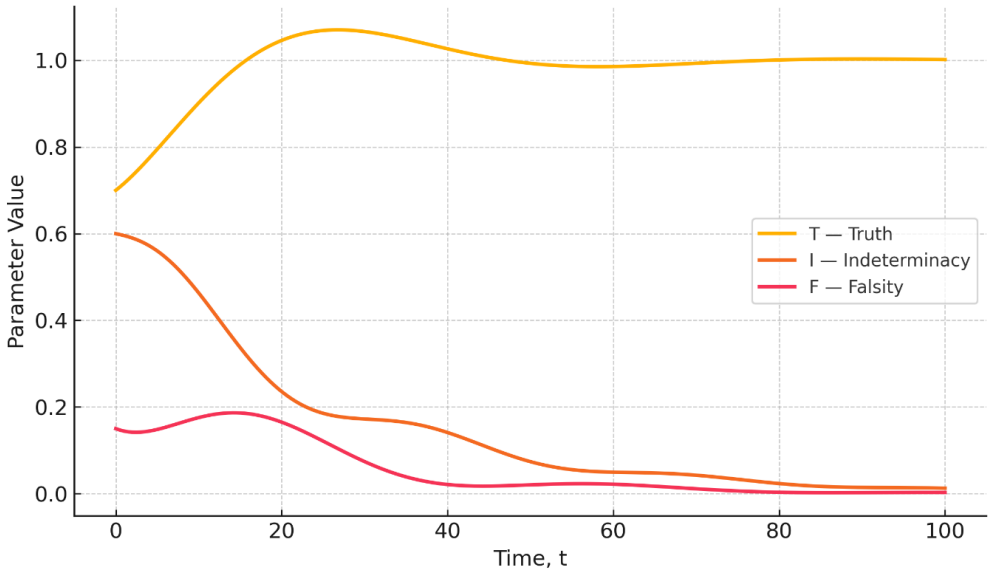


Figure 8. Dynamics of Neutrosophic Parameters T–I–F During Adaptation.

https://doi.org/10.31436/iiumej.v27i2.4021

The data show that the proposed NF-MRAC outperforms all compared algorithms in terms of accuracy, settling speed, and energy efficiency. Its advantages are especially evident in the presence of noise and structural uncertainty, where neutrosophic logic provides cognitive stability – a reduction of indeterminacy I and falsity F during the adaptation process. Figure 8 shows the typical dynamics of the change in the neutrosophic parameter triad T, I, F during network training.

Interpretation of the graph:

- At the initial stage ($t < 20$), high indeterminacy $I \approx 0.6$ dominates due to noise and weight instability.
- As training progresses, T increases, I decreases, and F tends to zero, reflecting the system's cognitive stabilization.
- By the time steady-state behavior is achieved ($t \approx 80$), the relationship $T \to 1, I \to 0.1, F \to 0.05$ is observed, which corresponds to the convergence of reasoning toward reliable knowledge.

Thus, neutrosophic analysis allows one to interpret the adaptation process as the evolution of model reliability over time, which serves as a unique cognitive metric for evaluating intelligent control systems.

7. DISCUSSION OF RESULTS

The obtained results confirm that the proposed neutrosophic-fuzzy MRAC effectively addresses the research problem of adaptive control of nonlinear dynamic systems under deep uncertainty. The observed 59% reduction in mean-square tracking error and nearly 1.8× faster convergence, compared to classical linear MRAC, are not merely quantitative improvements but reflect fundamental differences in the adaptation mechanism of the proposed controller. The reduction in tracking error can be attributed to the localized approximation capability of the B-spline-based neuro-fuzzy structure combined with the explicit representation of uncertainty through neutrosophic logic. Unlike classical MRAC, where parameter adaptation is driven by a global error signal, the proposed controller adjusts local submodels by considering not only the magnitude of the tracking error but also the associated degrees of truth, indeterminacy, and falsity. This mechanism suppresses the influence of unreliable or contradictory data, resulting in more stable parameter updates and improved approximation accuracy in regions of incomplete system knowledge.

The significant decrease in overshoot and oscillatory behavior is primarily explained by the smoothness properties of B-spline basis functions and the distributed nature of the control action. By allocating control effort across multiple localized submodels, abrupt parameter variations during transient processes are avoided. This leads to smoother state trajectories and prevents excessive control actions, which are typical for linear adaptive controllers operating under rapidly changing conditions. The introduction of neutrosophic principles provides an additional level of robustness that conventional neuro-fuzzy approaches cannot achieve. While classical fuzzy systems rely on a single degree of membership, the neutrosophic representation explicitly distinguishes between reliable information, uncertainty, and contradiction. This distinction is particularly important in practical control scenarios involving sensor noise, parameter drift, and partial observability, and directly contributes to the improved stability and robustness observed in the simulations.

From a computational perspective, the submodel decomposition strategy plays a critical role in enabling real-time applicability. By reducing the effective dimensionality of the approximation problem, the proposed controller achieves a 30–40% reduction in computational complexity compared to classical neuro-fuzzy models, without sacrificing control quality. This finding highlights the practical relevance of the proposed approach for embedded implementations.

The discussion of the results demonstrates that the proposed neutrosophic-fuzzy MRAC not only outperforms classical controllers numerically but also introduces a qualitatively different adaptive control paradigm. By integrating B-spline approximation, neuro-fuzzy learning, and neutrosophic uncertainty modeling, the controller provides a robust, interpretable, and computationally efficient solution to the problem of adaptive control of nonlinear systems under deep uncertainty. These findings directly support the research objective and confirm the scientific contribution of the proposed method. Future research may focus on extending the proposed framework to hybrid architectures that combine neutrosophic-fuzzy MRAC with deep learning techniques for online structural optimization, and on experimental validation on embedded hardware platforms.

8. LIMITATIONS OF THE PROPOSED STUDY

Despite the demonstrated effectiveness of the proposed neuro-neutrosophic MRAC, several limitations of the present study should be acknowledged. First, the validation of the proposed control strategy is currently limited to numerical simulations. Although the simulation results confirm improved robustness, stability, and energy efficiency under nonlinear dynamics and uncertainty, experimental verification on physical hardware platforms, such as digital signal processors or microcontroller-based control systems, has not yet been performed. Hardware implementation may introduce additional constraints related to quantization effects, computational delays, and real-time execution limits. Second, the stability analysis provided in this study focuses on local Lyapunov stability in the vicinity of the operating point. While this is sufficient for a wide class of practical adaptive control applications, global stability guarantees for strongly nonlinear systems with rapidly varying parameters remain an open research challenge. Third, although the proposed submodel decomposition strategy significantly reduces computational complexity, the selection of the number of B-spline basis functions and neutrosophic parameters still requires careful tuning. In systems with extremely high dimensionality, further optimization of the network structure may be necessary to ensure scalability and real-time performance.

These limitations do not diminish the practical value of the proposed approach but instead define clear directions for future research, including experimental validation on embedded hardware platforms, automated parameter optimization, and extension toward global stability analysis.

9. CONCLUSIONS

In this paper, a neuro-neutrosophic Model Reference Adaptive Control (NF–MRAC) framework based on B-spline neuro-fuzzy networks has been developed to address the problem of adaptive control of nonlinear dynamic systems operating under deep uncertainty. The proposed approach is specifically designed for situations where classical linear MRAC and conventional intelligent controllers exhibit degraded performance due to nonlinearities, parametric variability, external disturbances, and incomplete or contradictory state information. The main results of the study demonstrate that integrating neutrosophic logic into

the adaptive control loop significantly enhances robustness and stability. Comparative simulation analysis with a classical linear MRAC shows that the proposed NF–MRAC reduces the mean-square tracking error by approximately 59%, decreases overshoot by more than 3 times, shortens the transient response time by nearly 1.8 times, and lowers control energy consumption by approximately 18%. The use of B-spline basis functions ensures smooth control actions and localized approximation of nonlinear dynamics, while the proposed submodel decomposition strategy effectively mitigates the curse of dimensionality and reduces computational complexity. The core scientific contribution of this work lies in the formulation of a hybrid adaptive control architecture that explicitly incorporates neutrosophic uncertainty representation, i.e., truth, indeterminacy, and falsity, into an MRAC framework. Unlike traditional neuro-fuzzy and deep learning controllers, the proposed approach provides an interpretable, cognitively robust control mechanism capable of distinguishing among reliable, incomplete, and contradictory information. This makes the controller particularly suitable for complex nonlinear systems where epistemic uncertainty plays a dominant role. Despite the achieved improvements, the present study has several limitations. The validation of the proposed controller is currently restricted to numerical simulations, and only local Lyapunov stability has been analyzed. In addition, although the submodel decomposition significantly reduces the computational load, selecting B-spline parameters and neutrosophic coefficients may require careful tuning in high-dimensional systems. Future research will focus on the experimental implementation of the proposed NF–MRAC on embedded hardware platforms, such as digital signal processors and microcontroller-based control systems, to evaluate real-time performance, quantization effects, and computational constraints. Further extensions may include automated optimization of network parameters, investigation of global stability properties, and integration with advanced predictive and learning-based control strategies. Overall, the proposed neuro-neutrosophic MRAC introduces a qualitatively new adaptive control paradigm that explicitly models uncertainty within the control loop, providing a robust, interpretable, and computationally efficient solution for nonlinear systems operating under deep uncertainty.

REFERENCES

[1] Chen, J., Li, X., & Zhang, Y. (2020). Robust adaptive control for nonlinear aircraft system with uncertainties. *Applied Sciences, 10*(12), 4270. https://doi.org/10.3390/app10124270

[2] Rohrs, C. E., Valavani, L., Athans, M., & Stein, G. (1985). Robustness of continuous-time adaptive control algorithms in the presence of unmodeled dynamics. *IEEE Transactions on Automatic Control, 30*(9), 881–889. https://doi.org/10.1109/TAC.1985.1104070

[3] Smarandache F (2016). Neutrosophic Logic – A Generalization of the Intuitionistic Fuzzy Logic. SSRN Working Paper, January 25, 2016. https://doi.org/10.2139/ssrn.2721587

[4] Smarandache F (2018). Plithogeny, Plithogenic Set, Logic, Probability, and Statistics. arXiv preprint, August 12, 2018. https://arxiv.org/abs/1808.03948

[5] O. Porubay, "Multiscale analysis of wavelet transformation as a solution to the problem of compression of information flows," Proceedings of the 2016 International Conference on Information Science and Communications Technologies (ICISCT), p. 7777410, 2016. https://doi.org/10.1109/ICISCT.2016.7777410

[6] O. Porubay, I. Siddikov, G. Alimova, D. Umurzakova, and T. Abdullaev, “Adaptive Nonlinear Control of Electric Power Facilities Using a Synergetic Approach”, J Robot Control (JRC), vol. 6, no. 5, pp. 2380–2388, Oct. 2025. DOI: https://doi.org/10.18196/jrc.v6i5.27969

[7] U. Dilnoza Maxamadjonovna, "Neuro-fuzzy Control Algorithm of Dynamic Objects with Uncertainty of a Priori Information," 2020 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, 2020, pp. 1-4, doi: 10.1109/ICISCT50599.2020.9351462

[8] S. I. Xakimovich and U. D. Maxamadjonovna, "Neuro-fuzzy Adaptive Control system for Discrete Dynamic Objects," 2019 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, 2019, pp. 1-6, doi: 10.1109/ICISCT47635.2019.9012027.

[9] Z. A. Al-Dabbagh and S. W. Shneen, "Neuro-Fuzzy Controller for a Non-Linear Power Electronic DC-DC Boost Converters", J Robot Control (JRC), vol. 5, no. 5, pp. 1479–1491, Aug. 2024, doi: https://doi.org/10.18196/jrc.v5i5.22690

[10] M. Lazareva et al., "Optimization of operation modes of renewable energy facilities to provide energy for agriculture," E3S Web of Conferences, vol. 538, p. 01028, 2024. https://doi.org/10.1051/e3sconf/202453801028

[11] Oksana Porubay, Isamiddin Siddikov, and Dilnoza Umurzakova, "Intelligent control of energy system operating modes based on neuro-analytical and neutrosophic models under conditions of uncertainty", Neutrosophic Sets Syst., vol. 97, pp. 512–534, Mar. 2026, Accessed: Oct. 29, 2025. doi: 10.5281/zenodo.17420112.

[12] O. Porubay, I. Siddikov, G. Nashvandova, and G. Alimova, "Synthesis of a control system for a two-mass electromechanical object," AIP Conference Proceedings, vol. 3045, no. 1, p. 030080, 2024. https://doi.org/10.1063/5.0197280

[13] T. Abdullayev and A. Xoitqulov, "Development of a mathematical model of a temperature calibrator," AIP Conf. Proc., vol. 3045, no. 1, p. 030090, Mar. 2024, doi: https://doi.org/10.1063/5.0197324

[14] Heba Rashad, & Mai Mohamed. (2021). Neutrosophic Theory and Its Application in Various Queueing Models: Case Studies. Neutrosophic Sets and Systems, 42, 117-135.

[15] S. I. Xakimovich and U. D. Maxamadjonovna, "Synthesis of Adaptive Control Systems of a Multidimensional Discrete Dynamic Object with a Forecasting Models," 2019 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, 2019, pp. 1-5, doi: 10.1109/ICISCT47635.2019.9012033.

[16] O. Porubay and I. Siddikov, "Algorithms for optimization of operation modes of electric power systems under conditions of information uncertainty," Proceedings of the International Conference on Information Science and Communications Technologies (ICISCT), pp. 320–325, 2024. https://doi.org/10.1109/ICISCT64202.2024.10957429

[17] Abdel-Basset, M., & Mohamed, M. (2021). Multicriteria group decision making based on neutrosophic analytic hierarchy process: Suggested modifications. Neutrosophic Sets and Systems, 43, 247-254. Retrieved from https://digitalrepository.unm.edu/cgi/viewcontent.cgi?article=1848&context=nss_journal

[18] Chotikunnan, P., Chotikunnan, R., Nirapai, A., Wongkamhang, A., Imura, P., & Sangworasil, M. (2023). Optimizing membership function tuning for fuzzy control of robotic manipulators using PID-driven data techniques. *Journal of Robotics and Control (JRC)*, 4(2), 128–140. https://doi.org/10.18196/jrc.v4i2.18108

[19] Deli, Irfan; Vakkas Ulucay; and Zeynep Baser. "Neutrosophic Inference Systems Using Takagi-Sugeno-Kang Model and Its Application." Neutrosophic Sets and Systems 88, 1 (2025). https://digitalrepository.unm.edu/nss_journal/vol88/iss1/68

https://doi.org/10.31436/iiumej.v27i2.4134

A Comprehensive Framework of Ant Colony Optimization for Optimal Control of Upper Extremity Rehabilitation Robot

M. KAMRAN JOYO[1,*], IZANOORDINA AHMAD[1], KUSHSAIRY ABDUL KADIR[1], TALHA AHMED KHAN[2], NOAMAN M. NOAMAN[3]

[1]*Electronics Engineering Department, Universiti Kuala Lumpur British Malaysian Institute, 53100, Malaysia*
[2]*Cyber Security and Technological Convergence, Malaysian Institute of Information Technology (MIIT), Universiti Kuala Lumpur, Kuala Lumpur, Malaysia*
[3]*College of Engineering, University of Technology, Bahrain, Kingdom of Bahrain*

**Corresponding author: M. Kamran Joyo (muhammadkamran@unikl.edu.my)*

(Received: 14 January 2026; Accepted: 21 April 2026; Published online: 10 May 2026)

ABSTRACT: Passive rehabilitation of the upper limb requires precise joint positioning to facilitate motor recovery and ensure patient safety. For a two-degree-of-freedom (DOF) elbow robot, the primary challenge is unwanted disturbances arising from human coupling and voluntary and involuntary forces applied by the subject. This study addresses this issue by proposing a comprehensive control framework based on a nature-inspired meta-heuristic approach. The robot dynamics were obtained using the Lagrangian formulation, which was then combined with a realistic actuator model, yielding a closed-loop transfer function that describes system behavior. A two-degree-of-freedom (2-DOF) PID controller was used as the position controller. To determine the optimal PID parameters, an ant colony optimization (ACO) algorithm is used with an appropriate performance index to select the PID parameter values. The resulting ACO-based PID controller showed significant improvement over the conventional Ziegler–Nichols (Z-N) approach, as it could reject significant patient-induced forces, maintain the joint trajectory with the reference, and remain stable even under perturbations. Simulations were performed, confirming that the optimized controller delivers smooth, precise movements, reduces tracking error, and enhances the safety of passive rehabilitation procedures.

ABSTRAK: Pemulihan secara pasif di bahagian anggota atas pesakit memerlukan kedudukan sendi yang tepat bagi memudahkan pemulihan motor pesakit daripada bahaya atau cedera. Manakala pada robot siku dua darjah kebebasan (DOF), cabaran utama ialah mengelakkan gangguan yang tidak diingini oleh manusia dan daya sukarela atau tidak sukarela yang dikenakan oleh subjek. Kajian ini mencadangkan rangka kerja kawalan komprehensif berdasarkan pendekatan meta-heuristik yang diilham oleh alam semula jadi. Dinamik robot diwujudkan dengan menggunakan rumusan Lagrangian, diikuti dengan gabungan model penggerak realistik, agar dapat menghasilkan fungsi pemindahan gelung tertutup yang boleh menggambarkan tingkah laku sistem. Pengawal PID 2-DOF pula digunakan sebagai pengawal kedudukan. Bagi mencapai parameter optimum PID, algoritma Pengoptimuman Koloni-Semut (ACO) digunakan dengan indeks prestasi yang sesuai agar dapat memilih PID paling optimum. Pengawal PID berasaskan ACO yang terhasil mengatasi pendekatan Ziegler Nichols (Z-N) konvensional, kerana ia mempunyai keupayaan menolak daya teraruh pesakit yang ketara sambil mengekalkan trajektori bersama rujukan dan kekal stabil walaupun ketika gangguan. Simulasi yang dijalankan dalam kajian ini mengesahkan bahawa pengawal yang

https://doi.org/10.31436/iiumej.v27i2.4134

dioptimumkan berjaya mencapai pergerakan yang lebih lancar, tepat, seterusnya mengurangkan ralat pengesanan serta mempromosi keselamatan prosedur pemulihan pasif.

KEYWORDS: *Ant Colony Optimization, Rehabilitation Robot, PID, Ziegler Nichols*

1. INTRODUCTION

Stroke is the most common cause of long-term disability, and about half of the patients are left with upper limb motor impairment, which significantly impairs their quality of life [1, 2]. The process of recovery depends on intensive, repetitive, and task-specific exercises, which must begin in the early stages of the illness, during the acute and subacute stages [3]. Among rehabilitation modalities, passive range-of-motion exercises constitute a foundational therapeutic approach, particularly for patients with severe motor deficiencies who are unable to voluntarily initiate limb movements. These exercises prevent musculoskeletal problems, reduce spasticity, and facilitate the restoration of the functionality of disabled limbs [4]. Post-stroke therapy has been transformed by robotic rehabilitation systems, which provide specific and extensive repetitive movements with a significant reduction in the workload of therapists, allowing prolonged training sessions, and evidence demonstrates that extensive interventions significantly improve motor recovery [5].

Robot-assisted therapy offers advantages similar to those of conventional physiotherapy, providing consistent repeatability and performance tracking [6, 7]. Furthermore, combining passive range-of-motion exercises with visual and cognitive feedback mechanisms enhances motivation and engagement in therapy. Studies using functional near-infrared spectroscopy support the idea that robot-assisted therapy promotes neuroplasticity, thereby activating motor cortical regions and enhancing neural connections involved in functional recovery. The efficacy of robotic rehabilitation largely depends on the characteristics of the motion control employed within the system. In passive rehabilitation systems, the robot must track predefined trajectories with high precision, at the same time compensating for the mechanical properties and reactive forces exerted by the patient's limb [8].

Modern rehabilitation robots are employed in diverse configurations, including end-effector and exoskeleton-based structures, each presenting distinctive control challenges [9, 10]. These systems must accurately track predefined paths and, at the same time, handle variable human loads caused by heterogeneity in limb mass, involuntary contractions, and varying muscular flexibility [11]. The interaction between humans and robots is dynamic, leading to significant uncertainty in controller design. Such exercises prevent musculoskeletal problems, reduce spasticity, and facilitate the restoration of function in disabled limbs, thereby underscoring the need for robust control strategies [12]. The safety factors in physical HRI are the most important because excessive forces applied to soft and musculoskeletal tissues can cause injury, especially when applied for prolonged durations. However, small forces can be dangerous because of continuous contact and cyclic motion [9]. These human-coupled disturbances undermine the accuracy of position tracking, potentially leading to suboptimal therapeutic outcomes or patient discomfort.

Classical control methods for rehabilitation robots typically employ proportional-integral-derivative (PID) controllers tuned using heuristic methods such as Ziegler-Nichols. Despite the simplicity of PID control and its broad scope, heuristic-based parameterization is often unable to support multi-objective rehabilitation requirements, such as achieving rapid convergence, reducing overshoot, and maximizing disturbance rejection [13, 14].

https://doi.org/10.31436/iiumej.v27i2.4134

Existing studies have explored metaheuristic optimization techniques, which are recognized as effective methods for solving complex controller tuning problems. These include ant colony optimization (ACO), which has proven especially effective for finding optimal PID parameters by simulating ant colony foraging dynamics. The ACO exhaustively searches the parameter space by means of probabilistic exploration and exploitation to effectively identify near-optimal solutions [15]. Despite this potential, existing ACO-based rehabilitation control studies often suffer from limited comparative evaluations across different performance indices [16]. Furthermore, unlike previous studies that predominantly adapt discrete pheromone models to continuous tuning, this study employs ACO using a Gaussian kernel probability density function. This facilitates the exploration of search parameters in continuous space.

The contributions of this study include a dynamic and linearised model of a 2-DOF rehabilitation exoskeleton integrated with an actuator model. Second, a 2-DOF-PID architecture was applied to decouple reference tracking and disturbance in the presence of human-induced torques. Third, an ACO framework for a continuous domain was employed to tune the controller parameters using multiple integral performance indices. Fourth, a disturbance model was introduced, and robustness was assessed against sustained involuntary human-applied torques.

The remainder of this paper is organized as follows: Section 2 presents the dynamic modeling and controller design methodology. Section 3 presents the simulation results and optimization analysis; Section 4 provides a comprehensive comparative evaluation and discussion of the findings; and Section 5 concludes the paper with recommendations for future research directions.

2. MATERIALS AND METHODS

2.1. Overview of the RAX-1 Exoskeleton System

RAX-1 is an exoskeleton device developed for upper limb rehabilitation. The system is engineered to operate in coordination with the human forelimb, mimicking natural movement sequences to rehabilitate neuromuscular function. The human forelimb comprises nine degrees of freedom (DOFs), excluding the finger joints. In this study, the focus is on the elbow (humeroulnar) joint, which mainly involves flexion and extension movements and provides rotation of the forearm in its interaction with the humeroradial and proximal radioulnar joints.

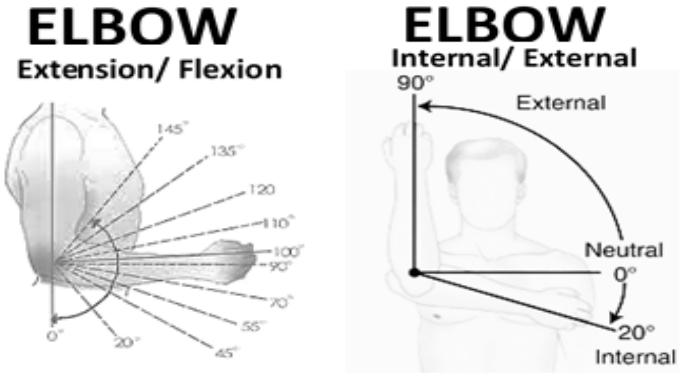


Figure 1. Elbow Extension/ Flexion and Elbow Internal/ External Rotation

Figure 1 depicts the principal motions of the elbow joint complex, and Table 1 lists the corresponding ranges of motion in healthy subjects [17]. These motion parameters were implemented as reference trajectories within the RAX-1 exoskeleton. The overall RAX-1

https://doi.org/10.31436/iiumej.v27i2.4134

configuration supports five DOFs, enabling exercises for the shoulder, elbow, and wrist. However, the shoulder and wrist joints were not actuated and were excluded from the control model, whereas only the elbow joints were actively controlled. A 3D model of the robot manipulator is illustrated in Figure 2.

Table 1. Range of Motion

Limb	Therapeutic exercise	ROM of Limb
Elbow	Extension/ Flexion	0° - 145°
	Internal/ External Rotation	-20° - 90°

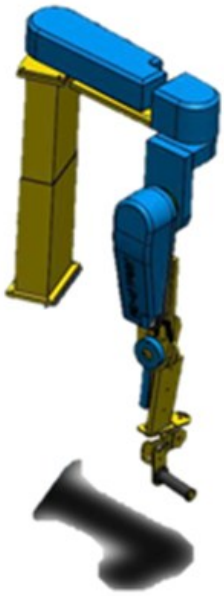


Figure 2. 3D CAD Model of RAX-1

2.2. Modeling using the Euler–Lagrange Approach

The dynamic model of the robotic manipulator was derived using the Euler–Lagrange formulation, which relates the joint positions and velocities to the system's kinetic and potential energies. This approach generalizes Newtonian mechanics to systems with position-dependent constraints. The Lagrangian L is defined as:

$$L = K - P \tag{1}$$

K and P denote the kinetic and potential energies, respectively. The generic mathematical expression that represents the motion of the jth joint is written as

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_j}\right) - \frac{\partial L}{\partial q_j} = \tau_j \tag{2}$$

where q_j represents the generalized coordinate, and τ_j is the required joint torque. The nonlinear dynamics derived above are linearized around the equilibrium point defined by the model initial conditions, i.e., $(q = 0, \dot{q} = 0)$ yielding a linear state-space representation $\dot{x} = Ax + Bu(t) + B_{hd}\tau_{hd}(t), y = Cx + Du(t)$, where $x = [q\ \dot{q}]^T$ suitable for control design. Since both actuator and human torque act on the same physical joint, the disturbance distribution matrix is $B_{hd} = B$. The human disturbance is defined as two input external torque vector $\tau_{hd}(t) = [\tau_{hd_1} \quad \tau_{hd_2}]^T$ representing the involuntary or resistive human interaction acting about the flexion/extension and forearm pronation/supination axes. The human

https://doi.org/10.31436/iiumej.v27i2.4134

disturbance torque was modeled as constant external load applied at each joint, with magnitudes selected as a fixed percentage of the actuator rated torque. These terms are defined as

$$\tau_{hd_1}(t) = \begin{cases} 0 & for\ t < t_d \\ \tau_{d_1} & for\ t > t_d \end{cases} \quad (3)$$

$$\tau_{hd_2}(t) = \begin{cases} 0 & for\ t < t_d \\ \tau_{d_2} & for\ t > t_d \end{cases} \quad (4)$$

where τ_{d_1} and τ_{d_2} represents the abrupt, sustained nature of involuntary muscle activation, spastic, and resistive behavior during passive rehabilitation rather than short-duration impulsive disturbances. The linearized model expresses the dynamics of a system as the relationship among velocity, position, and torque, which forms the basis for controller implementation.

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -30.351 & 0 & 0.0397 & 0 \\ 0 & 0 & 0 & 1 \\ 0.1685 & 0 & -30.023 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \\ 6.8916 & 0.0028 \\ 0 & 0 \\ 0.0028 & 6.7582 \end{bmatrix},$$

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad D = [0 \quad 0 \quad 0 \quad 0] \quad (5)$$

The exoskeleton joint is actuated by a DC motor that converts electrical input into rotational mechanical torque. Mechanical rehabilitation prefers DC motors owing to their low inertia, high torque density, and ability to control speed. The internal motor model is defined as follows:

$$V_a = L_a \frac{di_a}{dt} + R_a i_a + K_b \omega \quad (6)$$

$$T_m = K_\tau i_a = J \frac{d\omega}{dt} + B_m \omega + T_L \quad (7)$$

J is the inertia of the rotor in the above formula; B_m is the viscous damping coefficient; K_τ is the torque constant; K_b is the back EMF constant, and L_a, R_a are the armature inductance and resistance, respectively. In the case of $L_a \ll R_a$, the inductance term is very small and can be omitted, and the model is reduced to a first-order approximation. The motor is connected to a harmonic gear reduction unit to boost the torque output. Table 2 [18] provides detailed parameters.

Table 2. Motor Parameters

Parameters	Values
L_a (Henry)	2.6400e-04
R_a (ohms)	0.363
K_m (Torque Constant)	0.0705
J (Kg.m2/s2)	3.1212e-05
K_f (Nms)	8.66301e-03
K_b (back emf constant)	8.66301e-03
n (Gear Ratio)	75

It is important to acknowledge the limitations of the developed model. First, the robot dynamics were linearised around the equilibrium point defined by the model's initial conditions. Second, the human disturbance torque was modeled as a constant, sustained load

applied at each joint. Although this effectively simulates spasticity or a patient's persistent resistive force, it does not capture the full complexity of time-varying or impulsive interactions.

2.3. Control Algorithm

2.3.1. Two-Degree-of-Freedom PID Controller

A 2-DOF proportional–integral–derivative (PID) controller is incorporated to provide accurate tracking of the reference and to reject all disturbances. The controller continuously computes the difference between the desired and actual positions and adjusts the control signal accordingly [26]. The controller is mathematically expressed as

$$u(t) = K_p(b.r - y) + K_i(r - y)s^{-1} + K_d \frac{N}{1+Ns^{-1}}(c.r - y) \quad (8)$$

where K_p, K_i, K_d are proportional, integral, and derivative gains. In order to improve the flexibility and transient performance, the setpoint weighting parameters b and c are incorporated for the independent tuning of the proportional and derivative actions. MATLAB's 2-DOF PID block is utilized for implementation, allowing the adjustment of K_p, K_i, K_d, b, c, and filter coefficient N. This arrangement enables independent tuning of transient and steady-state responses for passive rehabilitation control.

2.3.2. Ant Colony Optimization for Continuous Domains

Ant colony optimization (ACO) is a population-based metaheuristic inspired by the foraging behavior of ant colonies, in which solutions are generated through the guided exploration of pheromone trails. To ensure continuous optimization, the pheromone model used by ACO employs a Gaussian kernel probability density function (PDF), which allows ants to sample continuous decision spaces instead of discrete solution elements [19]. The Gaussian kernel PDF is a weighted sum of the Gaussian function, which is defined as

$$G^i(x) = \sum_{l=1}^{k} \Omega_l g_l^i(x) = \sum_{l=1}^{k} \Omega_l \frac{1}{\sigma_l^i \sqrt{2\pi}} e^{-\left(\frac{(x-\mu_l^i)^2}{2\sigma_l^{i\,2}}\right)} \quad (9)$$

In the above formula, Ω is the weight vector of individual Gaussian functions, μ is the mean vector, and σ is the standard deviation. Solutions that are in an archive of size k parameterize this kernel, which stores the best solutions found during the search process. The ongoing ACO comprises three key elements:

Construction of Solution: Each ant constructs a solution through n construction steps for n decision variables. At Step i, the ant samples a variable via a two-stage process: first selecting a Gaussian function from the kernel with probability proportional to its weight.

$$P_l = \frac{\omega_l}{\sum_{r=1}^{k} \omega_r} \quad (10)$$

Then, samples were drawn from the selected Gaussian distribution. The standard deviation at the construction step i is computed as:

$$\sigma_l^i = \xi \sum_{e=1}^{k} \frac{\left|s_e^i - s_l^i\right|}{k-1} \quad (11)$$

where ξ is the pheromone evaporation parameter that controls convergence speed: larger values promote exploration while smaller values favor exploitation.

Pheromone Update: The solution archive maintains k high-quality solutions that define the PDF. After each iteration, newly generated solutions are added to the archive, and an equal number of the worst-performing solutions are removed, preserving the archive size and ensuring that the PDF reflects promising regions of the search space.

Daemon Actions: This optional component updates the best-found solution and implements termination criteria, typically based on the maximum iterations or convergence thresholds. The ACO parameters and values used in this study are presented in Table 3.

Table 3. ACO Parameters

Parameters	Values
Archive Size	20
Number of iterations	50
Sample Size	10
$\boldsymbol{q}$ (Selection pressure)	0.5
$\boldsymbol{\zeta}$ (Deviation distance ratio)	1

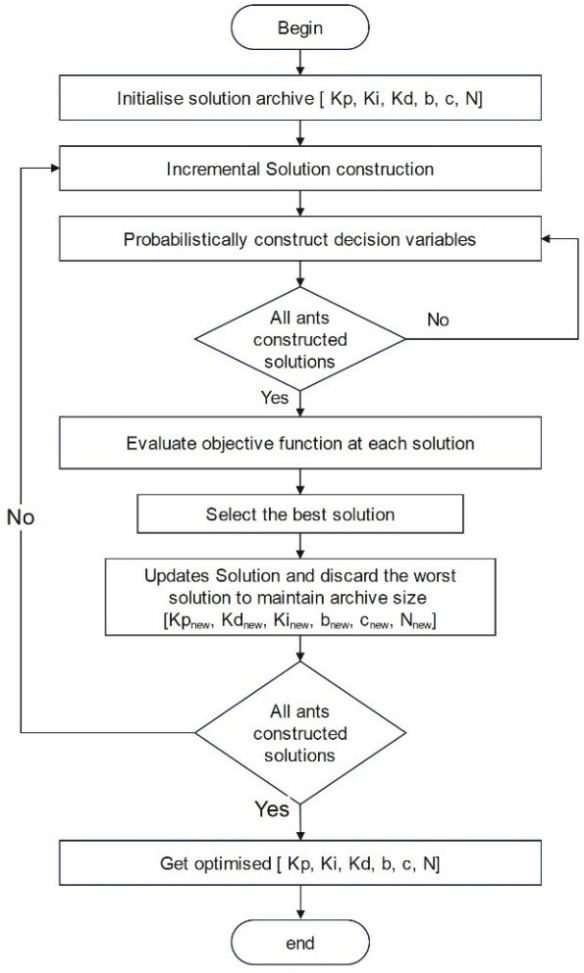


Figure 3. ACO metaheuristic framework for continuous optimization

The system was simplified to a linearized subsystem representing the elbow joint. A dedicated control loop equipped with saturation limits was used to regulate the robot's motion. A step input was applied to the elbow-joint model to assess the system response and evaluate the controller's overall performance. Figure 4 shows a schematic of the proposed control approach for utilizing the ACO algorithm for the elbow joint.

https://doi.org/10.31436/iiumej.v27i2.4134

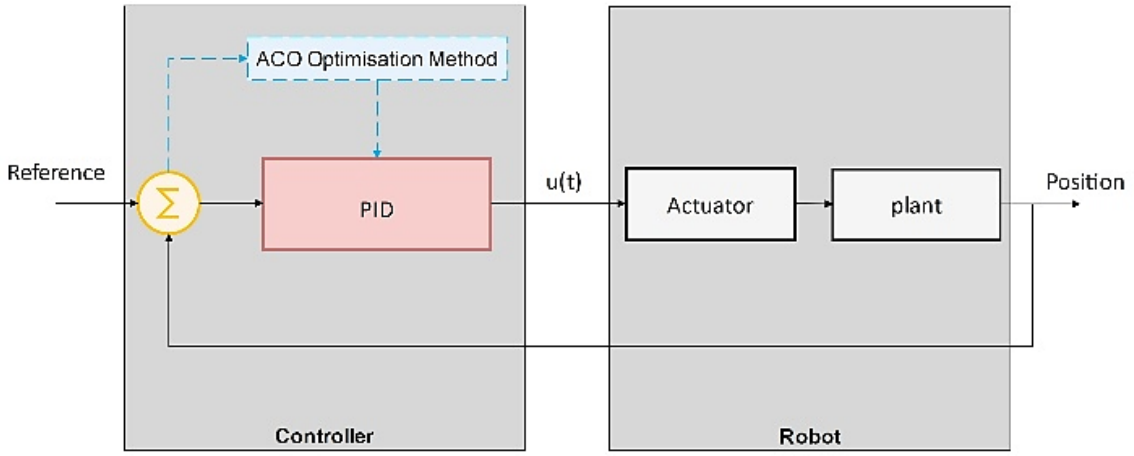


Figure 3. ACO-based proposed control structure

The effectiveness of model-based control systems depends on the selection of an appropriate performance index that encapsulates both the design criteria and system limitations. The performance index (objective or cost function) provides a measurable benchmark for optimization algorithms to minimize or maximize. A well-chosen performance index affects not only the algorithm's convergence behavior but also significantly influences the quality of the resulting controller parameters and, ultimately, the system's practical performance.

This study offers a comparative analysis of four integral-based performance indices to determine the most appropriate performance index for the upper-limb rehabilitation robotic system under study. The performance indices are Integral Squared Error (ISE), Integral Time Squared Error (ITSE), Integral Absolute Error (IAE), and Integral Time Absolute Error (ITAE). These performance indices are selected for their effectiveness in control systems for reference tracking and disturbance rejection. These performance measures have varying influences on the control elements. For example, the ISE and IAE aim to reduce the cumulative tracking error, whereas the time-weighted indices of ISE and ITAE apply a more severe punishment for extended deviations from the intended path. This difference plays a pivotal role in rehabilitation settings, where errors in reference tracking may impair patients' muscles. The objective here is to identify the performance index that can most effectively address the major goals, including correcting motion tracking, restricting overshoot, and improving disturbance rejection, to control unforeseen dynamics and forces exerted by the patient. The mathematical formulas for these functions are explained below.

$$ISE = \int e^2 dt \tag{12}$$

$$IAE = \int |e| dt \tag{13}$$

$$ITSE = \int t.e^2 dt \tag{14}$$

$$ITAE = \int t.|e| dt \tag{15}$$

The controller tuning process accounts for robustness to ensure stable closed-loop behavior in the presence of uncertainties, such as external disturbances. This is achieved using the maximum sensitivity function as a measure of robustness. It is defined as

$$M_s = \max_{\omega} |S(j\omega)| = \max_{\omega} \frac{1}{|1 - C_y(j\omega)P(j\omega)|} \tag{16}$$

where $S(j\omega)$ denotes the sensitivity function and $|S(j\omega)| \leq M_s$. The sensitivity function indicates how feedback affects the system output in response to disturbances. Disturbances are amplified when $|S(j\omega)| > 1$ and are attenuated when $|S(j\omega)| < 1$. Therefore, lower Ms values indicate greater robustness, as the system becomes less sensitive to uncertainties and external perturbations. The values of M_s ranging from 1.2 to 2.0 provides reasonable robustness [20].

3. RESULTS

An extensive set of simulation experiments was conducted in MATLAB to evaluate the effectiveness of the proposed control algorithm. A comparative study was conducted based on parameter analysis of the PID controller optimized using ACO and the conventional Z-N method. Ziegler-Nichols was selected as the baseline for comparison, largely because it is one of the most widely recognized classical approaches in the literature and offers an established standard. Several studies in rehabilitation systems have employed Z-N as their reference standard, enabling comparison. More importantly, Z-N is heuristic and does not optimize for specific integral-error criteria or disturbance rejection; it serves as an appropriate performance baseline against which the advantages of the ACO-based optimization method can be demonstrated. Table 4 lists the PID parameter ranges, and Table 5 presents the PID parameters obtained using ACO and Z-N. The analysis was narrowed to the motions of the elbow: extension and flexion, with internal and external rotation of the forearm.

Table 4. Parametric Constraints used for ACO and Z-N

Controller	Parameters	Elbow (E/F)	Elbow (I/E)
Ant colony	K_p	$0.01 < K_p < 10$	$0.01 < K_p < 10$
	K_i	$0.01 < K_i < 10$	$0.01 < K_i < 10$
	K_d	$0.01 < K_d < 10$	$0.01 < K_d < 10$
	b	$0.01 < b < 1$	$0.01 < b < 1$
	c	$0.01 < c < 1$	$0.01 < c < 1$
	N	$1 < N < 200$	$1 < N < 200$
Ziegler Nichols	K_u	4.5	5.4
	T_u	2.1	2.1

Table 5. Optimised parameters

Controller	Performance Index	Kp	Ki	Kd	b	c	N
		Elbow Extension/ Flexion					
PID-ACO	ISE	**2**	**1.9305**	**1.2024**	**0.47021**	**1.469**	**77.2565**
	IAE	**1.9799**	**2**	**0.59161**	**0.44925**	**0.93431**	**78.0639**
	ITSE	**1.9971**	**1.5836**	**1.1014**	**0.4737**	**1.5400**	**89.6019**
	ITAE	**1.9535**	**1.6479**	**0.4565**	**0.4388**	**0.8867**	**80.1878**
Ziegler-Nichols	-	2.16	3.6	0.324	1	1	100
		Elbow Internal/ External Rotation					
PID-ACO	ISE	1.1615	0.0001	2	0.9988	1.5588	48.1113
	IAE	2	1.4005	1.3777	0.8391	1.7570	111.3369
	ITSE	1.9328	1.7273	2	0.9141	1.7241	52.3570
	ITAE	1.8621	1.2025	1.2285	0.8678	0.9470	75.7466
Ziegler-Nichols	-	3.24	3.0857	0.8505	1	1	100

Figures 5 to 10 illustrate the step responses of the elbow joint using the IAE, ITAE, ISE, and ITSE performance indices, along with the conventional Ziegler–Nichols controller, for both joints. Figure 5a and 5b illustrate a step response of the elbow joint using IAE for flexion/extension and internal/external rotation, respectively. Figure 6a and 6b illustrate a step

https://doi.org/10.31436/iiumej.v27i2.4134

response of the elbow joint using ISE for flexion/ extension and internal/ external rotation, respectively.

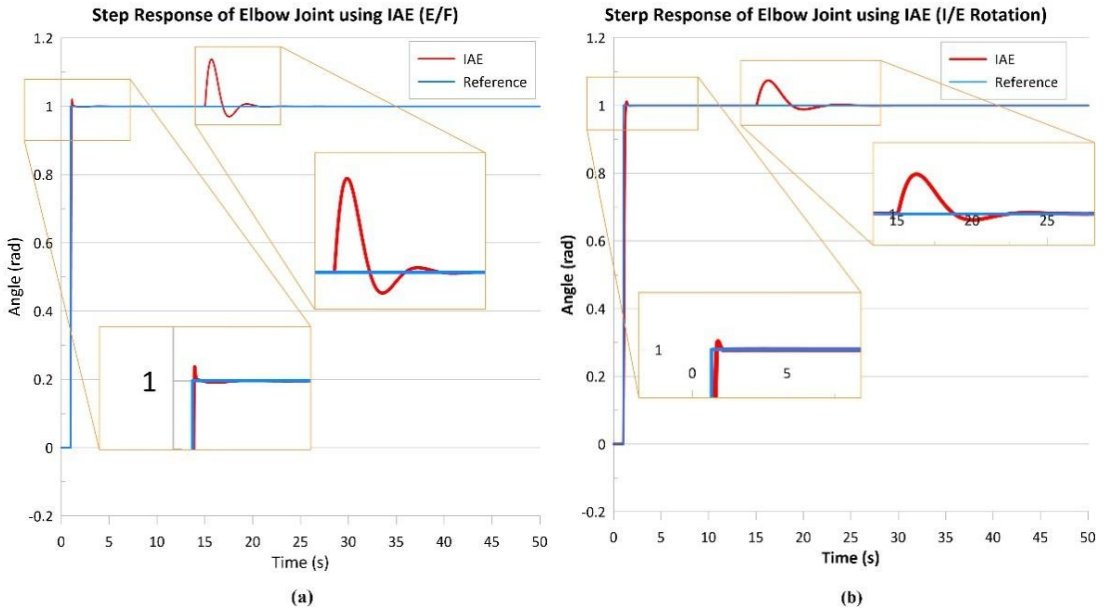


Figure 1. Step response of RAX-1 Elbow Joint using IAE optimized PID control (a) Extension/ Flexion, (b) Internal/ External Rotation

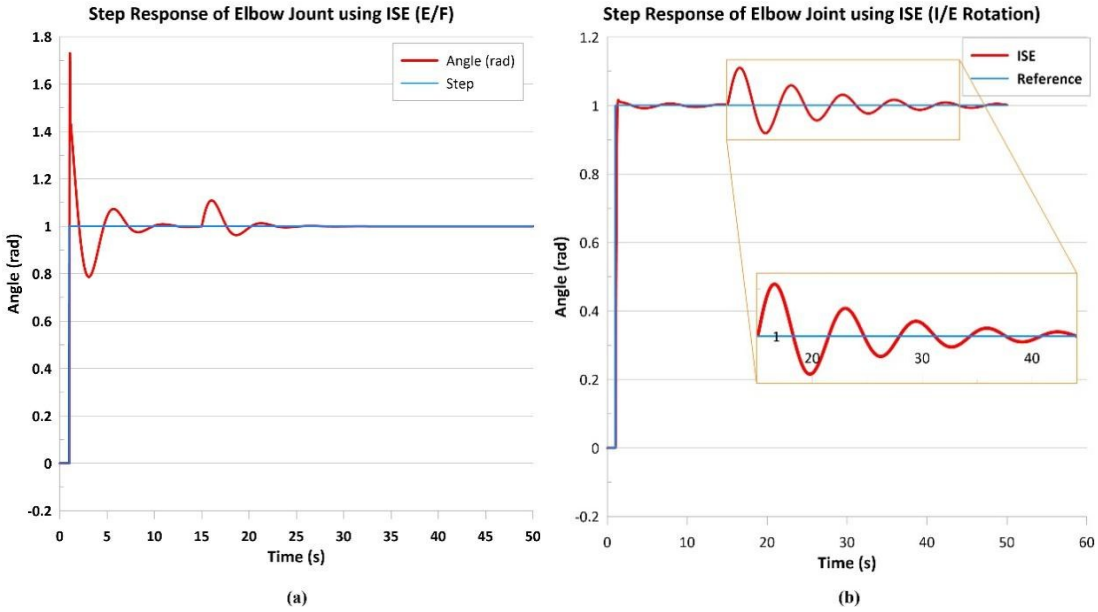


Figure 2. Step response of RAX-1 Elbow Joint using ISE optimized PID control (a) Extension/ Flexion, (b) Internal/ External Rotation

Figures 7a and 7b illustrate the step response of the elbow joint using ITAE for flexion/extension and internal/external rotation, respectively. Figure 8a and 8b illustrate a step response of the elbow joint using ITSE for flexion/ extension and internal/ external rotation, respectively.

https://doi.org/10.31436/iiumej.v27i2.4134

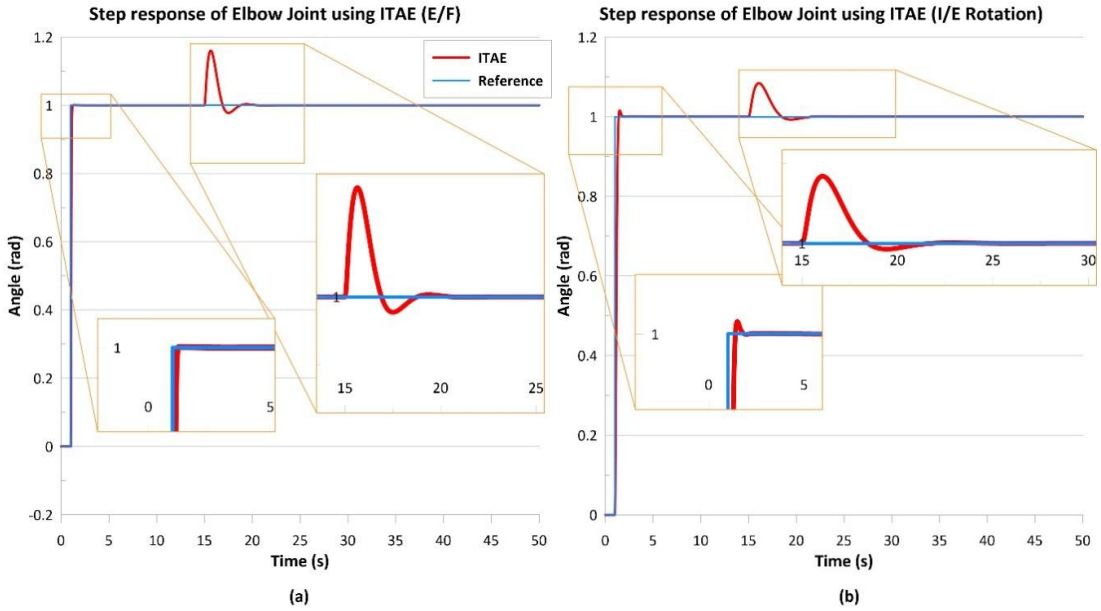


Figure 3. Step response of RAX-1 Elbow Joint using ITAE optimized PID control (a) Extension/ Flexion, (b) Internal/ External Rotation

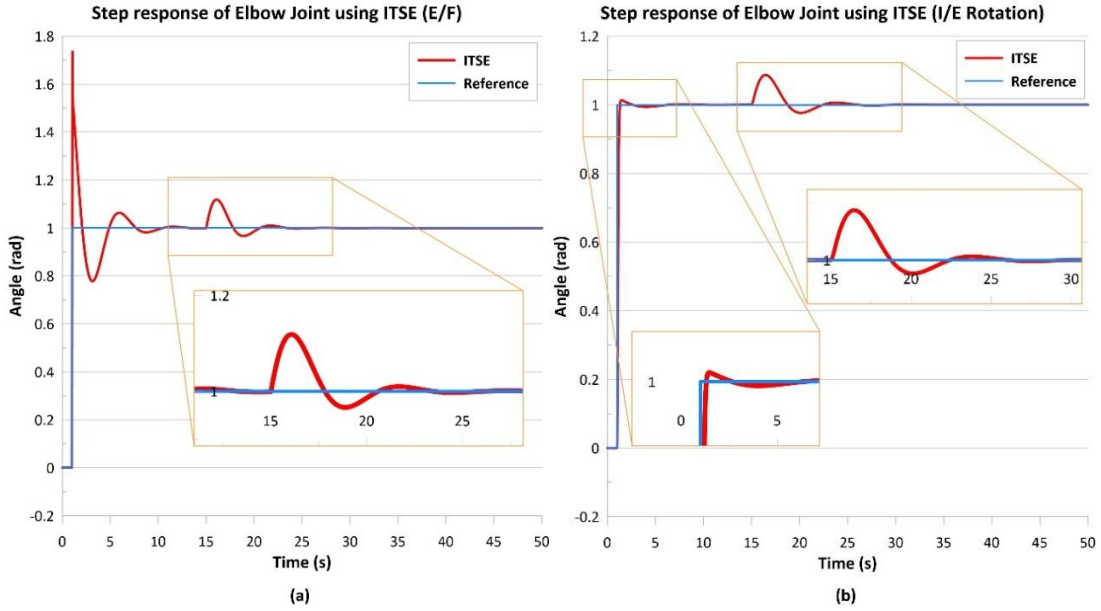


Figure 4. Step response of RAX-1 Elbow Joint using ITSE optimized PID control (a) Extension/ Flexion, (b) Internal/ External Rotation

Figures 9a and 9b illustrate the step response of the elbow joint using Z-N for flexion/extension and internal/external rotation, respectively.

https://doi.org/10.31436/iiumej.v27i2.4134

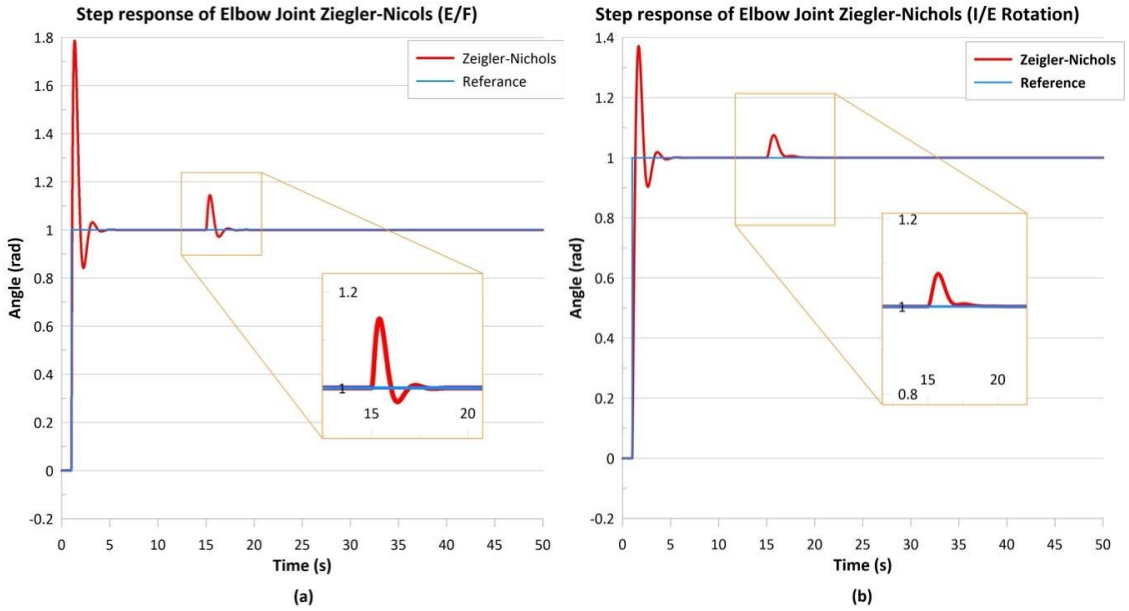


Figure 5. Step response of RAX-1 Elbow Joint Ziegler-Nichols (a) Extension/ Flexion, (b) Internal/ External Rotation

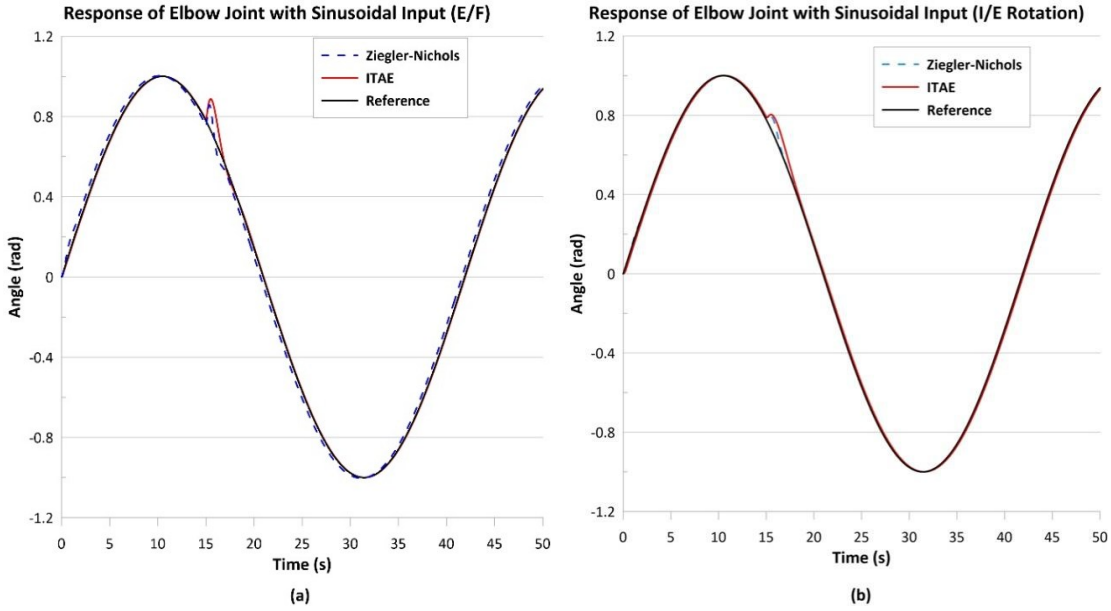


Figure 6. Sinusoidal tracking of RAX-1 Elbow Joint using ITAE optimized PID control and Ziegler-Nichols

Figures 10a and 10b show a comparison of the tracking performance of the ITAE-based ACO and Z-N tuned controllers for elbow joint flexion/extension and internal/external rotation when subjected to a sinusoidal reference trajectory.

4. DISCUSSION

Figure 5a and 5b illustrate the step response of the elbow joint under IAE-based ACO-PID control for flexion/extension and internal/external rotation, respectively. The ACO-tuned controller with IAE exhibits an immediate rise, minimal overshoot, and a stable settling region, demonstrating smooth tracking of the reference signal. The small, highlighted insets show the

minimal steady-state error and consistent transient performance throughout the simulation period, with effective rejection of the step disturbance introduced at 15s. This disturbance was intentionally introduced to represent voluntary or involuntary forces that a subject may apply during rehabilitation exercises. The controller maintains stable tracking despite this perturbation, demonstrating its ability to compensate for human-generated disturbances while preserving smooth and safe elbow joint motion. Figure 6a and 6b illustrate the step response of the elbow joint using ITAE for flexion/extension and internal/external rotation, respectively. The ACO-tuned controller with ITAE presents a similar stable response with slightly faster convergence than that with IAE. The ITAE criterion minimizes sustained oscillations by penalizing prolonged errors over time, resulting in a smoother trajectory closer to the reference. Figures 7a and 7b illustrate a step response of the elbow joint using ISE for flexion/extension and internal/external rotation, respectively. Figures 8a and 8b illustrate the elbow joint response for ITSE-tuned controllers. Both error-based indices maintain stable performance but exhibit higher oscillatory tendencies than the ITAE and IAE criteria. The ISE and ITSE-tuned controllers exhibit higher oscillatory behavior than the IAE and ITAE-based controllers. They place greater weight on early transient deviations, leading to more aggressive corrective actions that cause sudden overshoot, which is undesirable in any therapeutic exercise. In terms of minimizing the effects of disturbance, ITAE outperforms other functions. Figures 9a and 9b illustrate the response using Z-N for flexion/extension and internal/external rotation, respectively, showing a pronounced overshoot and a prolonged settling time, with noticeable oscillations around the reference position. This confirms that the Z–N method is less suitable for precise rehabilitation control tasks, where smooth motion and safety are critical.

Table 6. Performance Controller Parameters

Controller	Joint	O. S (%)	R.T (s)	S.T (s)	Ms
ACO-PID	Integral Squared Error				
	Extension/ Flexion	16.3%	0.0184	7.83	1.63
	Internal/ External Rotation	1.6%	0.129	16.5	1.23
	Integral Time Squared Error				
	Extension/ Flexion	10%	0.0182	6.2	1.53
	Internal/ External Rotation	1.3%	0.112	7.13	1.21
	Integral Absolute Error				
	Extension/ Flexion	2.08%	0.0569	0.131	1.34
	Internal/ External Rotation	3.8%	0.15	0.558	1.28
	Integral Time Absolute Error				
	Extension/ Flexion	Below 1%	0.0837	0.144	1.28
	Internal/ External Rotation	1.6%	0.234	0.359	1.32
Ziegler Nichols	Extension/ Flexion	78.6%	0.0654	2.47	1.53
	Internal/ External Rotation	40.2%	0.212	2.13	1.33

Table 6 presents the four performance indices for the PID parameters optimized by ACO, compared with the Z-N controller for elbow joint control. The findings highlight the distinct trade-offs of each performance index, providing insight into selecting the most appropriate performance criterion.

The Z-N-tuned controller exhibited a transient response characterized by a pronounced overshoot of 78.6% for elbow extension/flexion and 40.2% for internal/external rotation. Such excessive overshoot is not clinically acceptable in passive rehabilitation, since it may cause injury to the patient. In contrast, all ACO-tuned controllers drastically reduced the overshoot, demonstrating the fundamental advantage of the optimization approach. The Z-N-tuned controller showed a marginally faster rise time of 0.0653s for elbow extension/flexion, with a

settling time of 2.47s and a rise time of 0.234s for internal/external rotation, with a settling time of 2.13s. This represents the apparent speed at the cost of severe overshoot.

Among the four performance indices, the ITAE delivers the most balanced overall performance. For flexion/ extension movement, it achieves an overshoot of approximately 1%, combined with a settling time of 0.144s and a rise time of 1.28s. The maximum sensitivity index $M_s = 1.28$ falls within the accepted range, indicating good disturbance rejection. For internal/ external rotation, it achieves an overshoot of 1.6% with a settling time of 0.359s and a rise time of 0.234s. The $M_s = 1.32$ falls within the acceptable range. This combination of minimal overshoot, a reasonably quick settling time, and robustness makes ITAE a particularly compelling choice for applications where precise trajectory tracking is the predominant factor.

The IAE criterion also performs well, yielding a low overshoot of 2.08% for flexion/extension and the quick settling time of 0.131 s. This suggests that IAE is effective at quickly eliminating errors. However, its M_s value of 1.34 is slightly higher than that of ITAE, pointing to marginally lower robustness, although still within an acceptable range. For internal/external rotation, it exhibits an overshoot of 3.8%, a rise time of 0.15s, and a settling time of 0.558s. The value for M_s is 1.28, showing higher disturbance rejection.

Both quadratic indices produce less favorable transient responses. For the flexion/extension joint, ISE exhibits an overshoot of 16.3% with a rise time of 0.0184s and a settling time of 7.83 s, whereas ITSE exhibits a 10% overshoot with a rise time of 0.182s and a settling time of 6.2s, exhibiting noticeable oscillations. Aggressive corrective action that results from heavily penalizing large errors tends to induce a more nervous, oscillatory behavior, which is the type of response to avoid in a therapeutic context, where smooth, gentle motion is essential. The Ms value of 1.43 for ISE was achieved, whereas the Ms value for ITSE was 1.53, which is in the acceptable range. For internal/external rotation, ISE exhibits an overshoot of 1.6% with a rise time of 0.129s and a settling time of 16.5s, whereas ITSE exhibits a 1.3% overshoot with a rise time of 0.112s and a settling time of 7.13s, both indicating an oscillatory response.

Figure 10a and 10b show the tracking performance of the elbow joint flexion/extension and internal/external rotation when subjected to a sinusoidal reference trajectory. This test is particularly important for rehabilitation applications because sinusoidal motion closely resembles natural rhythmic movements commonly used during passive and assisted therapy sessions. The ACO-tuned controller, based on the ITAE criterion, closely follows the reference signal throughout the cycle, with only a slight deviation near the peak of the oscillation. The small variation settles rapidly, and the trajectory recalibrates with the nominal curve. The transient quickly stabilizes, and the trajectory moves to the intended position almost immediately. The Z-N controller, by contrast, exhibits a predictable phase lag, most pronounced around the ascending and descending zero-crossings. The delay is evidence of decreased responsiveness, even though the Z-N output eventually fully tracks the sinusoidal signal. The added perturbation makes the Z-N controller react significantly faster than the ACO-based approach. Nonetheless, this rapid response comes at the expense of precision, as the Z-N output exhibits steady-state error. In comparison, the ACO-tuned controller achieves superior tracking accuracy and smoothly returns to the reference trajectory, with a small residual error. This is an essential difference in rehabilitation, where rapid but inaccurate responses to situations can result in pain, whereas consistent, correct recovery can guarantee safer, more controlled joint movements.

5. CONCLUSION

This study presents the use of ant colony optimization to tune PID parameters for precise movement control of an elbow exoskeleton for use in the inactive form of forelimb rehabilitation. The proposed model is a synthesis of an Euler-Lagrange formulation that provides a 2-DOF PID structure for setpoint tracking and disturbance rejection, and a continuous-domain ACO algorithm that is employed to find an optimal solution.

The findings show that ACO significantly improves controller performance compared to the traditional Z–N approach. Of the four performance indices compared (ISE, ITSE, IAE, and ITAE), the ITAE criterion yielded the most appropriate performance. The overshoot was reduced from 78.6% (with Z-N) to approximately 1% for elbow flexion/ extension, achieving a settling time of 0.144s, and maintaining sustained disturbance rejection in the presence of simulated patient-induced forces. Similarly, for the internal/external rotation joint, the ITAE-optimized controller limited overshoot to 1.6%, settled in 0.359 s, and delivered an M_s of 1.32—well within the robust range. While the IAE criterion produced the fastest settling time, the ITAE-based controller offered the best overall combination of precision. Additionally, the sinusoidal trajectory was used as a reference, further demonstrating the effectiveness of the proposed method. The ACO-based controller proved useful for monitoring cyclic movements required during rehabilitation exercises, whereas the Z-N controller exhibited lag and steady-state errors.

Although the present study is limited to simulation analysis of the elbow joint on the RAX-1 exoskeleton, future research should focus on hardware implementation, multi-DOF control of the RAX-1 exoskeleton, and clinical tests involving patients with stroke to evaluate its actual performance, safety, and therapeutic effects.

ACKNOWLEDGEMENT

The author of this paper genuinely conveys their heartfelt appreciation for the comprehensive assistance provided by Universiti Kuala Lumpur in successfully concluding this project. This assistance comprised monetary support, the use of other facilities, such as the Intelligent Embedded Research Lab (IERL) research cluster, the experimental laboratory at the UniKL BMI, and the efforts of all contributors.

REFERENCES

[1] Lee SH, et al. (2020). Comparisons between end effector and exoskeleton rehabilitation robots regarding upper extremity function among chronic stroke patients with moderate to severe upper limb impairment. Scientific Reports, 10(1):1-8. https://doi.org/10.1038/s41598-020-58630-2

[2] Mehrholz J, Pollock A, Pohl M, Kugler J, Elsner B. (2020). Systematic review with network meta analysis of randomized controlled trials of robotic assisted arm training for improving activities of daily living and upper limb function after stroke. Journal of NeuroEngineering and Rehabilitation, 17(1):1-14. https://doi.org/10.1186/S12984-020-00715-0

[3] Ranzani R, et al. (2020). Neurocognitive robot assisted rehabilitation of hand function: a randomized control trial on motor recovery in subacute stroke. Journal of NeuroEngineering and Rehabilitation, 17(1):1-13. https://doi.org/10.1186/S12984-020-00746-7

[4] Dalla Gasperina S, Roveda L, Pedrocchi A, Braghin F, Gandolla M. (2021). Review on patient cooperative control strategies for upper limb rehabilitation exoskeletons. Frontiers in Robotics and AI, 8:745018. https://doi.org/10.3389/frobt.2021.745018

[5] Ai Q, Liu Z, Meng W, Liu Q, Xie SQ. (2023). Machine learning in robot assisted upper limb rehabilitation: a focused review. IEEE Transactions on Cognitive and Developmental Systems, 15(4):2053-2063. https://doi.org/10.1109/TCDS.2021.3098350

[6] Escarabajal RJ, Pulloquinga JL, Zamora-Ortiz P, Valera A, Mata V, Valles M. (2023). Imitation learning based system for the execution of self paced robotic assisted passive rehabilitation exercises. IEEE Robotics and Automation Letters, 8(7):4283-4290. https://doi.org/10.1109/LRA.2023.3281884

[7] Brown DF, Xie SQ. (2024). Effectiveness of intelligent control strategies in robot assisted rehabilitation: a systematic review. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 32:1828-1840. https://doi.org/10.1109/TNSRE.2024.3396065

[8] Brahmi B, Driscoll M, El Bojairami IK, Saad M, Brahmi A. (2021). Novel adaptive impedance control for exoskeleton robot for rehabilitation using a nonlinear time delay disturbance observer. ISA Transactions, 108:381-392. https://doi.org/10.1016/j.isatra.2020.08.036

[9] Ranzani R, Chiriatti G, Schwarz A, Devittori G, Gassert R, Lambercy O. (2023). An online method to monitor hand muscle tone during robot assisted rehabilitation. Frontiers in Robotics and AI, 10:1093124. https://doi.org/10.3389/frobt.2023.1093124

[10] Cao J, et al. (2024). Variable admittance control of high compatibility exoskeleton based on human robotic interaction force. Chinese Journal of Mechanical Engineering, 37(1):119. https://doi.org/10.1186/S10033-024-01113-6

[11] Yang R, Zheng J, Song R. (2022). Continuous mode adaptation for cable driven rehabilitation robot using reinforcement learning. Frontiers in Neurorobotics, 16:1068706. https://doi.org/10.3389/fnbot.2022.1068706

[12] Zhou J, Peng H, Zheng M, Wei Z, Fan T, Song R. (2024). Trajectory deformation based multi modal adaptive compliance control for a wearable lower limb rehabilitation robot. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 32:314-324. https://doi.org/10.1109/TNSRE.2023.3348332

[13] Wang Y, Held D, Erickson Z. (2022). Visual haptic reasoning: estimating contact forces by observing deformable object interactions. IEEE Robotics and Automation Letters, 7(4):11426-11433. https://doi.org/10.1109/LRA.2022.3199684

[14] Yang J, Su H, Li Z, Ao D, Song R. (2016). Adaptive control with a fuzzy tuner for cable based rehabilitation robot. International Journal of Control, Automation and Systems, 14(3):865-875. https://doi.org/10.1007/S12555-015-0049-4

[15] Poudel YK, Bhandari P. (2024). Control of the BLDC motor using ant colony optimization algorithm for tuning PID parameters. Archives of Advanced Engineering Science, 2(2):108-113. https://doi.org/10.47852/BONVIEWAAES32021184

[16] Mirrashid N, Alibeiki E, Rakhtala SM. (2022). Development and control of an upper limb rehabilitation robot via ant colony optimization PID and fuzzy PID controllers. International Journal of Engineering, Transactions B Applications, 35(8). https://doi.org/10.5829/IJE.2022.35.08B.04

[17] Biodex. (n.d.). System 3 and 4 upper extremity hemiparetic attachments operation manual. Retrieved from https://www.biodex.com

[18] Joyo MK, et al. (2019). Optimized proportional integral derivative controller for upper limb rehabilitation robot. Electronics, 8(8):826. https://doi.org/10.3390/electronics8080826

[19] Moradi B, Kargar A, Abazari S. (2022). Transient stability constrained optimal power flow solution using ant colony optimization for continuous domains. IET Generation Transmission and Distribution, 16(18):3734-3747. https://doi.org/10.1049/gtd2.12560

[20] Precup R. E., Preitl, S., and Korondi, P. (2007). Fuzzy controllers with maximum sensitivity for servosystems. IEEE Transactions on Industrial Electronics, 54(3):1298–1310. https://doi.org/10.1109/TIE.2007.893053

https://doi.org/10.31436/iiumej.v26i2.4136

Optimal and Robust Speed Control of Electric Vehicles using a Hybrid Ant Colony Optimization Tuned Sliding Mode Control

PARTHIBAN S, KRISHNA S*

School of Mechanical Engineering, Vellore Institute of Technology, Vellore Campus, India

**Corresponding author: krishna.s@vit.ac.in*

(Received: 15 January 2026; Accepted: 16 March 2026; Published online: 10 May 2026)

ABSTRACT: Electric vehicle (EV) speed control is significantly affected by nonlinear motor dynamics, external disturbances, and parameter uncertainties, which degrade tracking performance and energy efficiency. Conventional controllers often fail to maintain robustness under varying operating conditions. Although Sliding Mode Control (SMC) provides strong robustness, it suffers from chattering and requires proper tuning of control parameters. To address this, a hybrid Ant Colony Optimization-based Sliding Mode Controller (ACO-SMC) is proposed to optimally tune the SMC control parameters. The performance of the ACO-SMC controller is evaluated using the US06 and FTP-75 driving cycles in MATLAB Simulink. The proposed controller outperforms the ACO-tuned proportional-integral-derivative (ACO-PID), conventional SMC, and PID controllers. The ACO-SMC controller significantly improves stabilization and robustness, even in the presence of measurement noise, external disturbance, and varying vehicle parameters. The results reveal that the hybrid ACO-SMC controller effectively minimizes chattering and maintains accurate speed control, even during sudden load disturbance and reference speed variations across all examined driving cycles. The integral absolute error (IAE) and integral squared error (ISE) are used as performance indices, achieving their minimum values over the complete US06 and FTP-75 driving cycles, thereby confirming the high accuracy of the proposed approach.

ABSTRAK: Kawalan kelajuan kenderaan elektrik (EV) dipengaruhi secara signifikan oleh dinamik motor tidak linear, gangguan luaran, serta ketidakpastian parameter, seterusnya menjejaskan prestasi penjejakan dan kecekapan tenaga. Pengawal konvensional lazimnya gagal mengekalkan keteguhan di bawah operasi yang berubah-ubah. Walaupun Kawalan Mod Luncur (SMC) menawarkan keteguhan yang tinggi, kaedah ini mengalami masalah *chattering* dan memerlukan penalaan parameter kawalan yang teliti. Bagi mengatasi isu ini, satu pengawal hibrid berasaskan Pengoptimuman Koloni Semut (ACO-SMC) dicadangkan bagi menala parameter SMC secara optimum. Prestasi pengawal ACO-SMC dinilai menggunakan kitaran pemanduan US06 dan FTP-75 pada MATLAB Simulink. Pengawal yang dicadangkan menunjukkan prestasi yang lebih baik berbanding pengawal PID yang ditala ACO (ACO-PID), SMC konvensional, dan PID biasa. ACO-SMC meningkatkan kestabilan dan keteguhan secara signifikan walaupun dalam kehadiran hingar pengukuran, gangguan luaran, serta variasi parameter kenderaan. Dapatan kajian menunjukkan bahawa pengawal hibrid ini berupaya meminimumkan fenomena *chattering* dan mengekalkan kawalan kelajuan yang tepat, termasuk semasa gangguan beban mendadak dan variasi rujukan kelajuan merentasi semua kitaran pemanduan yang diuji. Indeks prestasi seperti ralat mutlak berkamiran (IAE) dan ralat kuasa dua berkamiran (ISE) mencapai nilai minimum sepanjang satu kitaran penuh bagi kedua-dua kitaran pemanduan US06 dan FTP-75, sekaligus mengesahkan ketepatan tinggi melalui kaedah yang dicadangkan.

https://doi.org/10.31436/iiumej.v26i2.4136

KEYWORDS: *Sliding Mode Control (SMC), Ant Colony Optimization (ACO), EV speed control, MATLAB Simulink, Drive cycle*

1. INTRODUCTION

EVs are rapidly becoming the preferred option for sustainable and environmentally responsible transportation worldwide. The propulsion system plays a crucial role in determining the overall EV performance. Researchers from industry and academia have worked to improve control systems used in electric vehicle propulsion [1]. The controller should ensure low tracking error and minimal energy use. However, EV systems are highly nonlinear and time-varying, and are affected by changing road conditions, motor behavior, and external disturbances. This makes designing a controller that can handle disturbances and uncertainty while using low control effort challenging [2].

Traditional PID controllers are often used in industrial applications because they are simple and require minimal tuning [3]. But, they perform poorly under diverse operating conditions and do not guarantee the necessary dynamic performance [4]. Various methods have been proposed to fine-tune PID controllers, with the Ziegler–Nichols method being a widely recognized classical approach. This technique often provides good performance across many practical control systems [5]. However, in some cases, it fails, resulting in a significant overshoot in the system response.

Classical PI and PID controllers struggle under dynamic load conditions, prompting the development of adaptive control methods such as fuzzy logic (FL), adaptive Neuro-Fuzzy Inference System (ANFIS), Artificial Neural Network (ANN), and Sliding Mode Control (SMC). ANN controllers have gained attention for their effective performance in motor control applications, particularly for speed regulation and torque ripple reduction [6]. An optimization strategy was proposed in [7] to improve EV energy efficiency by reducing battery charging costs. Moreover, combining ANN techniques with traditional PID controllers yields significantly improved control performance, mainly due to the self-tuning capabilities of ANN algorithms [8]. However, ANN controllers face challenges, including large data requirements and long training times.

The fuzzy control method, based on fuzzy sets, linguistic rules, and decision-making theory, is an adaptive approach used in many applications [9]. Although fuzzy control has yielded successful results, correctly establishing the controller's rule base remains an important issue. SMC has become a highly effective technique for managing nonlinear systems due to its inherent robustness, strong stability, and ability to suppress external disturbances and parameter uncertainties. It is widely used in power electronics applications [10-11]. SMC ensures a rapid dynamic response by forcing the system's state trajectory to reach and stay on a predefined surface called the sliding surface. As a variable structure control method, it offers finite-time convergence and resilience to modeling inaccuracies, making it a reliable choice for modern control system design [12].

As noted in [13], the effectiveness of sliding mode control depends heavily on the selection of parameters that define the sliding surface. SMC involves two control signals: one during reaching mode and one in sliding mode. According to [14], switching control (u_{sw}) is employed during reaching mode to bring the system state onto the sliding surface, while the equivalent control (u_{eq}) in sliding mode ensures stable system behavior. A major drawback of this method is its discontinuous behavior, which causes high-frequency switching, or chattering, once the sliding surface is reached. This switching can excite unmodeled system

dynamics, leading to instability, degraded performance, or damage. Several approaches exist to reduce chattering. One common method is replacing the discontinuous switching function with a smooth, continuous one near the sliding surface. Another is high-order sliding mode control [15], which reduces chattering by controlling the higher-order derivatives of the sliding variable. These methods typically lower control gain as the system approaches the sliding surface, allowing the equivalent control to dominate once the surface is reached.

Alternative intelligent control methods, such as integrating SMC with fuzzy logic [16] or using Particle Swarm Optimization (PSO) [17], have been explored to enhance performance. However, these methods have limitations: fuzzy logic relies on heuristic rules with limited adaptability, and PSO and Genetic Algorithms (GAs) often suffer from premature convergence and require careful tuning. In contrast, Ant Colony Optimization (ACO) is a powerful metaheuristic that offers better global search, faster convergence, and adaptability to changing environments. These advantages make ACO more suitable for optimizing controller parameters in nonlinear and uncertain systems [18]. However, most existing ACO-SMC studies are limited to idealized step-reference tracking and do not adequately address realistic electric-vehicle operating conditions, including standard driving cycles and implementation constraints.

To address these limitations, this study presents an improved ACO-SMC-based speed control strategy for electric vehicles by integrating the optimization capability of ACO with the robustness of SMC, thereby achieving accurate, stable, and robust speed control across diverse driving conditions. Unlike existing studies that primarily focus on step reference inputs, the proposed controller is evaluated under standard electric vehicle driving cycles such as FTP-75 and US06, which represent practical vehicle operation. In addition, the controller's performance is quantitatively analyzed with respect to speed-tracking accuracy, disturbance-rejection capability, control-signal smoothness, chattering amplitude, and control RMS. Furthermore, system constraints such as control delay and voltage limitations are incorporated to ensure realistic performance evaluation. The main contributions of this work are summarized below:

- An optimized ACO-SMC controller is developed for electric vehicle speed control and validated under FTP-75 and US06 driving cycles.
- Comparative performance evaluation is conducted against SMC, PID, and ACO-PID controllers using ISE and IAE performance indices.
- Robustness of the proposed controller is verified through set-point tracking, external disturbance rejection, noise suppression, and sensitivity analysis.
- Chattering reduction and control signal smoothness are quantitatively analyzed using chattering amplitude and RMS value.

The rest of this paper is structured as follows: Section 2 covers the mathematical model of an electric vehicle. Section 3 outlines the proposed ACO-SMC hybrid controller framework for EV speed control. Section 4 presents the results and discussions. Section 5 summarizes the conclusions.

2. MATHEMATICAL MODEL OF ELECTRIC VEHICLE

The mathematical model of an EV is essential for analyzing its dynamic behavior and designing effective control strategies. The electric vehicle system primarily consists of a battery unit, a controller unit, and an electric motor that transmits power to the drivetrain. The dynamic behavior of the EV can be divided into two primary components: vehicle dynamics and motor

https://doi.org/10.31436/iiumej.v26i2.4136

dynamics. Despite PMSM and BLDC motors being widely used in modern EVs, DC motors remain suitable for low-power electric vehicles such as electric buggies and electric golf carts due to their simple dynamics and ease of control [18-21]. Therefore, a DC motor-driven electric vehicle is adopted in this work to validate the proposed ACO-SMC controller.

2.1. Vehicle Dynamics

Modeling the electric vehicle involves analyzing the balance of forces acting on the vehicle during motion. The key forces considered include rolling resistance (F_r), aerodynamic drag (F_{ad}), gravitational force (F_g), and the vehicle acceleration force (F_a) as illustrated in Figure 1. The total traction force (F_t) acts on the electric vehicle is expressed in Equation 2.

$$F_t = F_a + F_r + F_{ad} + F_g \tag{1}$$

$$F_t = ma + \mu_r mgcos\theta + 0.5\rho C_d A_f v^2 + mgsin\theta \tag{2}$$

Where m is the mass of the electric vehicle, a is the change in velocity over time, μ_r is the rolling resistance coefficient, g is the acceleration due to gravity, ρ is the air density, C_d is the drag coefficient, A_f is the vehicle frontal area, v is the vehicle velocity, θ is the slope angle.

The net force F_t acting on the vehicle generates a load torque T_L that is applied to the drive motor and can be expressed as,

$$T_L = F_t \times \frac{r}{G} \tag{3}$$

Where, G is the gearing ratio and r is the EV tire radius.

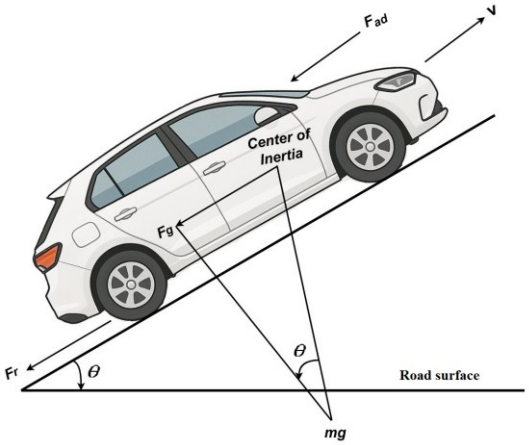


Figure 1. External forces acting on the EV

2.1. Motor dynamics

Considering a DC motor-driven electric vehicle, the motor dynamics can be modeled as,

$$\frac{d_i}{d_t} = \frac{V}{L_a+L_f} - \frac{(R_a+R_f)i}{L_a+L_f} - \frac{L_{af}i\omega}{La+L_f} \tag{4}$$

$$\frac{d\omega}{dt} = \frac{L_{af}i^2}{J} - \frac{B\omega}{J} - \frac{T_L}{J} \tag{5}$$

Where, ω is the motor angular speed, i is considered the field and armature current, L_f is the field winding inductance, L_a is the armature inductance, L_{af} is the mutual inductance, J is the motor moment of inertia, B is the viscous coefficient, V is the input voltage, T_L is the load torque, R_a is the armature resistance and R_f is the field winding resistance.

https://doi.org/10.31436/iiumej.v26i2.4136

The velocity to drive the vehicle is given by,

$$v = \omega \times \frac{r}{G} \tag{6}$$

The complete EV model, which combines both motor and vehicle dynamics, can be written as follows [20] and is implemented in Simulink, as illustrated in Figure 2. Table 1 describes the key parameters of the EV and the DC motor, along with their electrical specifications.

$$\frac{di}{d_t} = \frac{1}{(L_a + L_f)}\left[V - (R_{a+}R_f)i - i\omega L_{af}\right] \tag{7}$$

$$\frac{dw}{dt} = \frac{1}{\left[J+m\left(\frac{r}{G}\right)^2\right]}\left[L_{af}\, i^2 - B\omega - \frac{r}{G}\left(\mu_r mg + 0.5\rho A_f C_d v^2 + mg\, sin\theta\right)\right] \tag{8}$$

Equations 7 and 8 represent a nonlinear model that can be further converted into a state-space model.

$$\dot{x} = A(x) + B(x)\, u, y = C(x) \tag{9}$$

where

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} i \\ \omega \end{bmatrix} \tag{10}$$

$$A(x) = \begin{bmatrix} -\dfrac{R_a + R_f}{L_a + L_f}\, x_1 - \dfrac{L_{af}}{L_a + L_f}\, x_1 x_2 \\ \dfrac{1}{J + m\left(\dfrac{r^2}{G^2}\right)}\left\{L_{af} {x_1}^2 - B x_2 - \dfrac{r}{G}\left(\mu_r mg + 0.5\rho A_f C_d \dfrac{r^2}{G^2}\, {x_2}^2 + mg\, sin\theta\right)\right\} \end{bmatrix}$$

$$B(x) = \begin{bmatrix} \dfrac{1}{L_a + L_f} \\ 0 \end{bmatrix}, C(x) = x_2$$

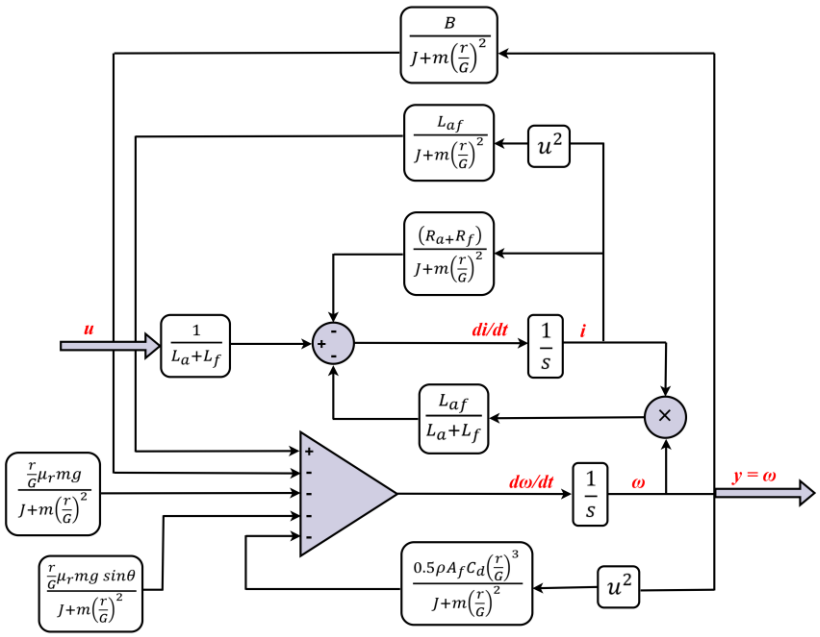


Figure 2. Overall structure of the EV system in Simulink

Table 1. EV specifications [20]

Parameters	Value	Parameters	Value
m	800 kg	$L_a + L_f$	6.00 mH
A_f	1.8 m^2	$R_a + R_f$	0.12 Ω
ρ	1.25 (kg/m^3)	J	0.05 kg m^2
C_d	0.3	L_{af}	1.76 mH
$c\theta$	0º	B	0.0002 N m s
G	11	i	78 A (250 max)
μ_r	0.015	V	48 V
r	0.25 m	ω	25 km/h

2.3. Inverter and Drive System Constraints

In practical electric vehicle systems, inverter efficiency, voltage limits, and control delays affect motor performance. In this study, these effects are considered by incorporating voltage saturation and system dynamics into the model. The control input voltage applied to the motor is limited using saturation bounds based on the rated motor voltage to ensure operation within the permissible range of the drive system. Additionally, the system dynamics inherently reflect the effects of inverter and actuator response delays. These constraints prevent the generation of unrealistic control signals and ensure safe motor operation. Thus, the simulation model provides a more realistic representation of the practical behavior of electric vehicle drive systems.

3. ACO-SMC Optimization Technique for EV Speed Improvement

The SMC controls the system's state trajectory toward a predefined sliding surface based on the system error, which is the deviation between the reference and the actual output. Once the system reaches this surface, it is forced to slide along it toward the equilibrium point. However, the conventional SMC controller suffers from excessive chattering and poor adaptability, which can directly affect the speed tracking performance and overall system stability. To overcome these challenges, this study uses an ACO-SMC-based controller that improves the system's adaptability and significantly reduces chattering. The proposed optimal SMC controller scheme for EV speed control is illustrated in Figure 3.

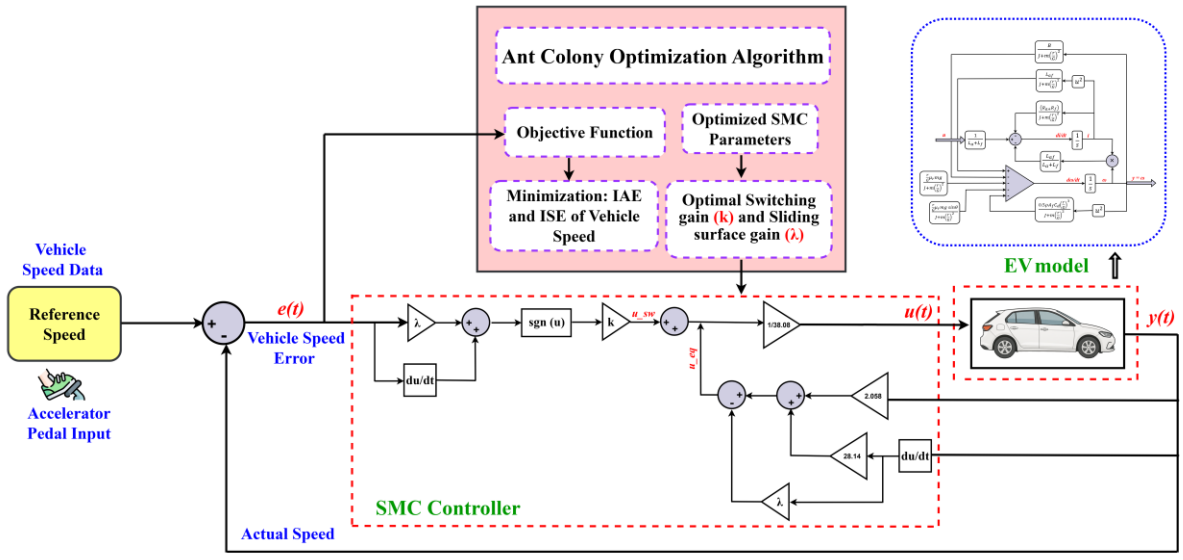


Figure 3. Proposed optimal SMC controller block diagram for EV speed control

https://doi.org/10.31436/iiumej.v26i2.4136

The control system integrates an ACO algorithm with the SMC to achieve robust and optimized EV speed regulation. The input is the reference speed from the driver's accelerator pedal, compared with the actual vehicle speed to generate the speed error e(t). This error signal is fed into the SMC controller, which computes the control signal $u(t)$ to operate the EV system. The performance of the SMC controller depends on the switching gain (K) and sliding surface gain (L). Instead of manual tuning, the ACO finds its optimal values by minimizing the IAE and ISE error metrics. Once these parameters are optimized, they are fed into the SMC controller to enhance speed accuracy and reduce chattering, ensuring adaptive EV speed control. The parameters of the ACO algorithm used for this study were selected based on standard recommended values from [18] and are listed in Table 2.

Table 2. ACO Parameters [18]

Parameters	Value
α	0.8
β	0.2
ρ	0.7
δ	0.01
No. of ants	30
No. of Iteration	30

The EV parameters in Table 1 are used in the simulation and controller design. The SMC controller is formulated using the state-space controllable form derived from Equation 9.

$$\ddot{\omega} + 28.14\dot{\omega} + 2.058\omega = 38.08v \tag{10}$$

Let us define the state variables as follows: $x_1 = \omega$, $x_2 = \dot{\omega}$ and take the control input as $u = v$. With these definitions, the system can be represented in state-space form using the controllable canonical structure:

$$\dot{x}_1 = x_2$$
$$\ddot{\omega} + 28.14\dot{\omega} + 2.058\omega = 38.08v \tag{11}$$

A sliding surface is then constructed by using the tracking error.

$$e(t) = \omega_{ref}(t) - \omega(t) \tag{12}$$

A first-order sliding surface is then formulated.

$$s(t) = \dot{e}(t) + Le(t)$$
$$s(t) = \dot{\omega}_{ref}(t) - \dot{\omega}(t) + L\left(\omega_{ref} - \omega\right) \tag{13}$$

where, ω_{ref} is the reference speed, ω is the actual speed, $e(t)$ is the error signal, and L is a positive sliding surface gain coefficient.

By using the Lyapunov function,

$$V(s) = \frac{1}{2}s^2 \tag{14}$$

To ensure system stability, the time derivative of the Lyapunov function presented in equation 14 must satisfy the condition $\dot{V}_{(s)} < 0$.

$$s\dot{s} < 0 \tag{15}$$

The derivative of the sliding mode function is

$$\dot{s} = \ddot{e} + L\ddot{e} \tag{16}$$

https://doi.org/10.31436/iiumej.v26i2.4136

$$\dot{s} = L\dot{e} + \ddot{\omega}_{r_{ef}} + 28.14\dot{\omega} + 2.058\omega - 38.08u \tag{17}$$

Thus, the derivative of the Lyapunov function $\dot{V}_{(s)}$ is

$$\dot{s}s = \left(L\dot{e} + \ddot{\omega}_{r_{ef}} + 28.14\dot{\omega} + 2.058\omega - 38.08u\right)s \tag{18}$$

To satisfy the condition $s\dot{s} < 0$, the control input is designed using the sliding mode control law given as,

$$u = \frac{1}{38.08}\left(2.058\omega + 28.14\dot{\omega} + \ddot{\omega}_{ref} + L\dot{e} + K\,sgn(s)\right) \tag{19}$$

The sliding mode controller was modified to reduce the chattering effect, as

$$u = \frac{1}{38.08}\left(2.058\omega + 28.14\dot{\omega} + \ddot{\omega}_{ref} + L\dot{e} + K\,\frac{s}{|s|+\delta}\right) \tag{20}$$

Where,

$$sign(s) = \begin{bmatrix} 1 & if\ s > 0 \\ 0 & if\ s = 0 \\ -1 & if\ s < 0 \end{bmatrix} \tag{21}$$

δ is the scalar tuning parameter to reduce the chattering effect, and K is the switching gain coefficient.

Substitute (20) for (15), then we get,

$$s\dot{s} = -k\,sgn(s) < 0 \tag{22}$$

ACO is a robust and adaptive metaheuristic algorithm inspired by the foraging behavior of real ant colonies. Ants communicate indirectly through pheromone deposition, allowing them to find optimal paths or solutions over time. In this algorithm, artificial ants explore possible paths from an initial state to a target solution, guided by a combination of pheromone trails and heuristic information. The process is iterative, and paths associated with better solutions receive higher pheromone deposits, making them more attractive to subsequent ants. The pheromone matrix, written as $\tau = \{\tau_{ij}\}$ is used to find the optimal solution. The initial state of the pheromone matrix is given by,

$$\tau_{ij} = \tau_0\,\psi_{(i,j)} \tag{23}$$

where, $\tau_0 > 0$. The probability $\{P_{ij}^A\}$ of selecting a node i at node j is given as

$$P_{ij}^A = \frac{[\tau_{ij}]^{\alpha}\cdot[\eta_{ij}]^{\beta}}{\Sigma_{(i,j)\in T^A}[\tau_{ij}(t)]^{\alpha}\cdot[\eta_{ij}]^{\beta}} \tag{24}$$

where T^A represents the path followed by an ant A at a given time, α and β are indeed constants that control the impact of pheromone values and the influence of the heuristic value on an ant's decision. The heuristic factor η_{ij} is given by

$$\eta_{ij} = \frac{1}{d_{ij}}, i = [L, K] \tag{25}$$

where d_{ij} is the distance between nodes i and j.

Using the SMC control law in equation 20, the parameters K and L are evaluated using the heuristic factor in equation 25. The quality of the pheromones $\Delta{\tau_{ij}}^A$ at each path is defined as

https://doi.org/10.31436/iiumej.v26i2.4136

$$\Delta\tau_{ij}{}^{A} = \begin{bmatrix} \frac{L^{min}}{L^{A}} \\ 0 \end{bmatrix} \quad (26)$$

where, L^A is the objective function value found by the ant A and L^{min} is the best solution carried out by the set of ants A until the current iteration.

The pheromone updating is,

$$\sum_k \Delta\tau^k = \Delta_\tau^{k=1} = \frac{\zeta f_b}{f_w} \quad (27)$$

Pheromone evaporation plays a crucial role in preventing the excessive accumulation of pheromone trails,

$$\tau_{ij}(t) = (1-\rho)\tau_{ij}(t-1) + \sum_{A=1}^{NA} \Delta\tau_{ij}^{A}(t) \quad (28)$$

where NA is the number of ants, while ρ is the evaporation rate.

A controller is considered optimal when its control parameters are tuned to minimize a predefined cost function. The objective function (K, L) is formulated to minimize the integral absolute error (IAE) and integral squared error (ISE) between the reference and the actual speed over a specified time duration.

$$\text{IAE} = \int_0^\infty |\text{e}(t)|\, \text{d}t \quad (29)$$

$$\text{ISE} = \int_0^\infty e^2(t)\, \text{d}t \quad (30)$$

4. Results and Discussion

The proposed ACO-SMC controller minimizes the objective functions defined in equations 29 and 30 to obtain the optimal control parameters for the SMC controller. To evaluate the controller's performance under realistic conditions, the FTP-75 and US06 driving cycles are used as reference speed profiles. The controller is tested under four main scenarios: set-point tracking, noise suppression, disturbance rejection, and sensitivity analysis to examine robustness, accuracy, and effectiveness.

4.1. Evaluating ACO-SMC for FTP-75 and US06 driving cycles

To evaluate the performance of the proposed ACO-SMC controller for an EV, the FTP-75 (cold stabilization phase) and US06 driving cycle were used as the reference speed profiles to represent urban and highway driving conditions, respectively. The SMC controller calculates the tracking error $e(t)$ as the difference between the reference speed and the actual speed. Initially, the sliding surface gain L = 5 and switching gain K = 10 for FTP-75, L = 5 and K = 10 for US06 were assigned for controller tuning. Afterward, they are tuned using the ACO algorithm for 30 iterations, yielding optimized gain values of L = 9.809 and K = 16.426 for FTP-75, and L = 9.096 and K = 29.42 for US06, based on the evaluation of performance metrics IAE and ISE.

4.1.1. Set-point tracking

The set-point tracking performance of PID, conventional SMC, ACO-PID, and the proposed ACO-SMC controllers was evaluated under both FTP-75 and US06 driving cycles. During the FTP-75 cycle, the PID, SMC, and ACO-PID controllers exhibit noticeable tracking errors, particularly during rapid speed transitions between 460 s and 510 s. In contrast, the

https://doi.org/10.31436/iiumej.v26i2.4136

proposed ACO-SMC controller closely follows the reference speed with minimal tracking error, as shown in Figures 4a and 4b. Similarly, under the aggressive US06 driving cycle, the proposed controller maintains superior tracking performance despite rapid speed variations between 146 and 186 seconds, as shown in Figures 5a and 5b. The control signal comparison is shown in Figures 6a and 6b for FTP-75 and US06, respectively. It is observed that the PID and conventional SMC controllers produce high-frequency oscillations and larger control variations due to poor tuning and chattering effects. The ACO-PID controller reduces these oscillations to some extent, whereas the proposed ACO-SMC controller produces the smoothest control signal with minimal oscillations.

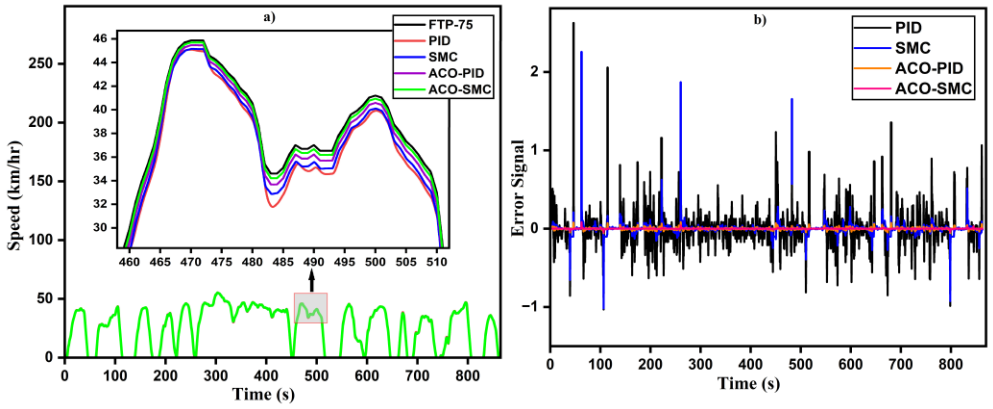


Figure 4. a) FTP-75 Speed tracking, b) FTP-75 Error analysis

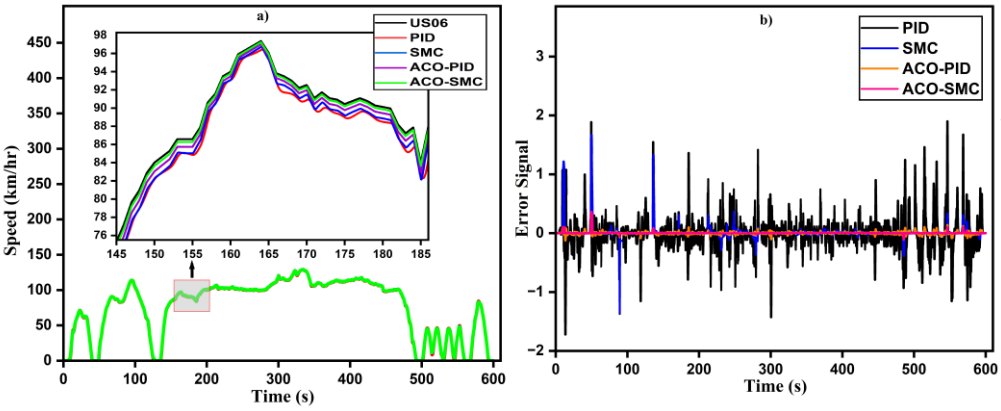


Figure 5. a) US06 Speed tracking, b) US06 Error analysis

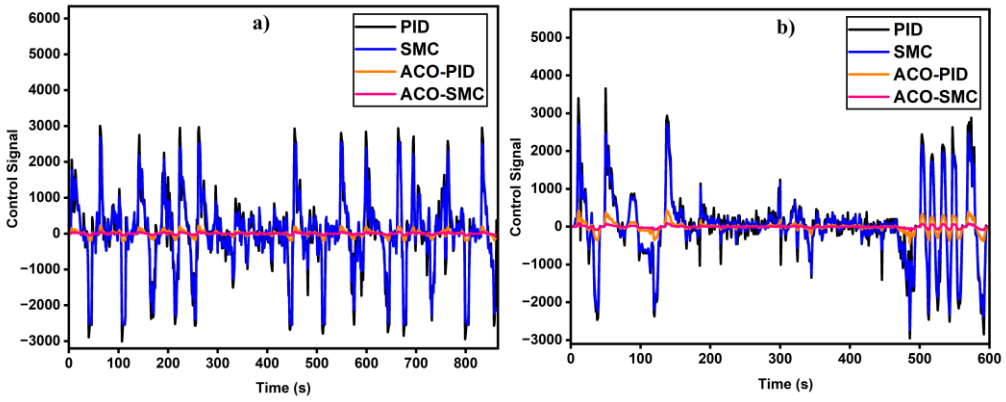


Figure 6. a) FTP-75 Control effort, b) US06 Control effort

Table 3. Performance metrics for FTP-75 and US06

	Performance Metrics	IAE	ISE	Control RMS	Chattering Amplitude
FTP-75	PID	10.03	1.594	2.85	1.92
	SMC	6.329	0.762	2.21	1.35
	ACO-PID	2.521	0.071	1.74	0.82
	ACO-SMC	1.525	0.029	1.12	0.38
US06	PID	12.26	2.135	3.94	2.85
	SMC	9.01	1.806	2.95	1.96
	ACO-PID	4.88	0.422	2.18	1.21
	ACO-SMC	3.25	0.152	1.48	0.62

Table 3 summarizes the performance metrics for the FTP-75 and US06 cycles. The proposed ACO-SMC controller achieves the lowest IAE and ISE values for both cycles, indicating improved tracking accuracy. Furthermore, the ACO-SMC control RMS and chattering amplitude values are significantly lower than those of PID, SMC, and ACO-PID, indicating smoother and more efficient control. Overall, the proposed ACO-SMC controller provides superior tracking performance, smoother control signal, and reduced chattering under both FTP-75 and US06 driving cycles.

4.1.2. Noise suppression

To evaluate the robustness of the ACO-SMC controller, measurement noise was introduced into the speed feedback signal to represent realistic sensor inaccuracies, with an amplitude range of -0.05 to +0.05 and a sampling time of 0.01 seconds. As shown in Figure 7, the PID, conventional SMC, and ACO-PID controllers exhibit noticeable speed fluctuations under noisy conditions. In contrast, the proposed ACO-SMC controller effectively suppresses noise-induced oscillations by maintaining a smoother, more stable speed response, demonstrating superior noise rejection.

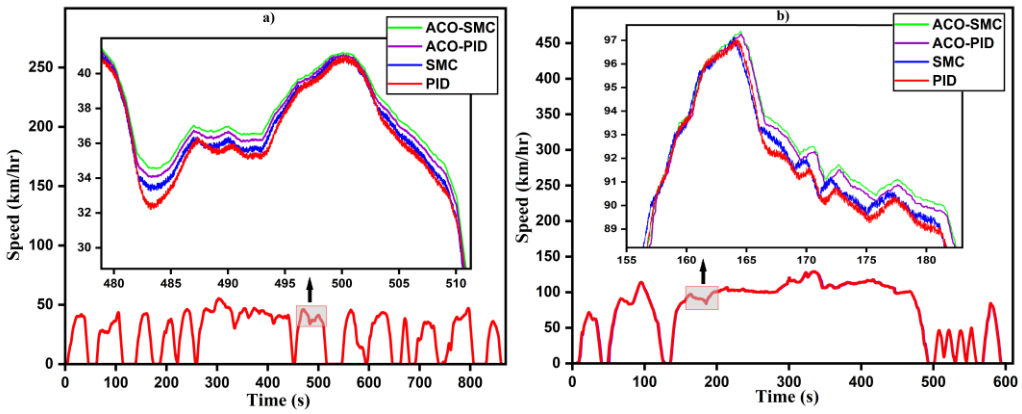


Figure 7. Performance of the controller in the presence of measurement noise for FTP-75 and US06

4.1.3. External disturbance rejection

To evaluate the effectiveness of the ACO-SMC controller, an external disturbance was introduced into the system as a step signal of 1km/h, representing a static disturbance at t = 303 s for the FTP-75 and t = 163 s for the US06 cycle, respectively. As shown in Figure 8, the PID, conventional SMC, and ACO-PID controllers exhibit noticeable deviations from the reference speed due to the external disturbance. In contrast, the proposed ACO-SMC controller effectively restores the vehicle speed to the reference speed with a shorter recovery time,

demonstrating superior robustness against external disturbances and better speed regulation performance.

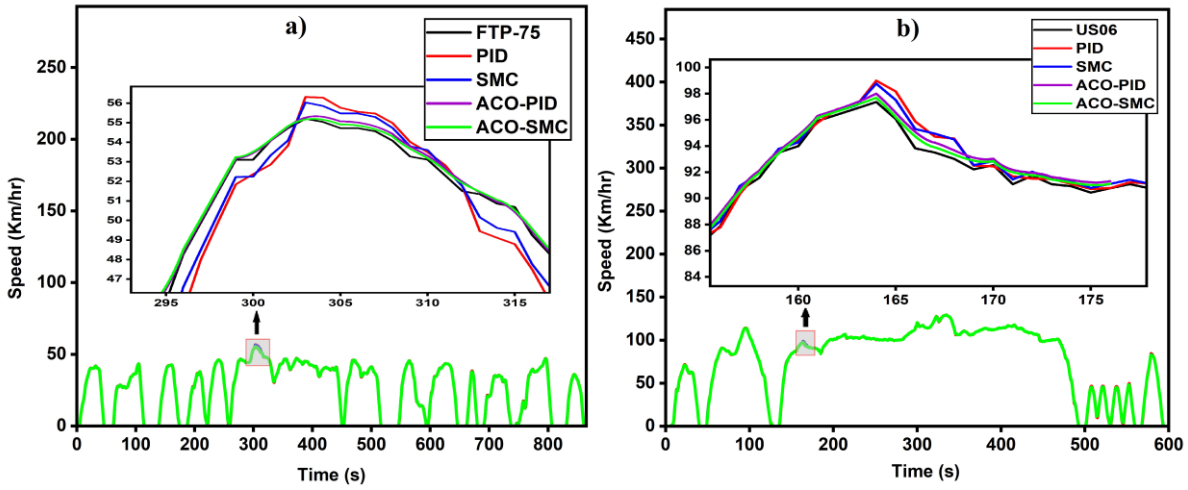


Figure 8. Performance of the controller under the influence of external disturbance for FTP-75 and US06

4.1.4. Sensitivity analysis

The robustness of the controllers is tested by varying key EV system parameters such as mass (m), drag coefficient (C_d), rolling resistance coefficient (μ_r), and EV tire radius (r). Figure 9 illustrates the robustness of the proposed controller under variations in EV system parameters, including a 20% increase in mass (m), 25% increase in tire radius (r), 20% decrease in drag coefficient(C_d), and a 30% increase in rolling resistance coefficient (μ_r). It is observed that the ACO-SMC controller completes full-power acceleration faster and stabilizes faster than the other controllers in the presence of uncertainties.

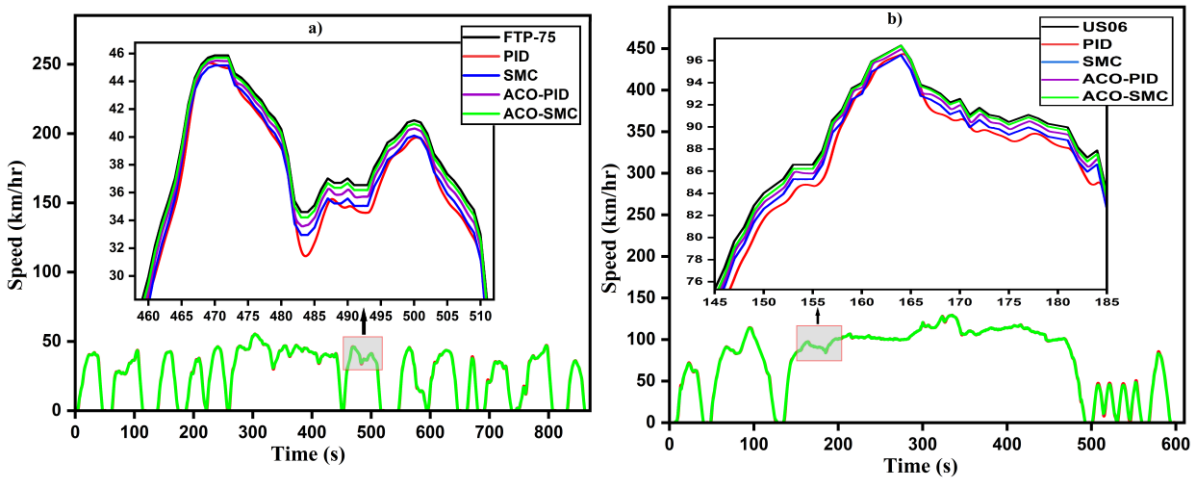


Figure 9. Robustness of the controller against the EV parameter variations for FTP-75 and US06

4.2. Frequency-Domain Analysis of ACO-SMC Controller

The sensitivity function represents the system's ability to achieve accurate set-point tracking, while the complementary sensitivity function indicates the robustness against external road disturbance noise. It is observed from Figure 10 that the sensitivity function must have

https://doi.org/10.31436/iiumej.v26i2.4136

smaller values at low frequencies, while the complementary sensitivity function in Figure 11 must have smaller values at higher frequencies.

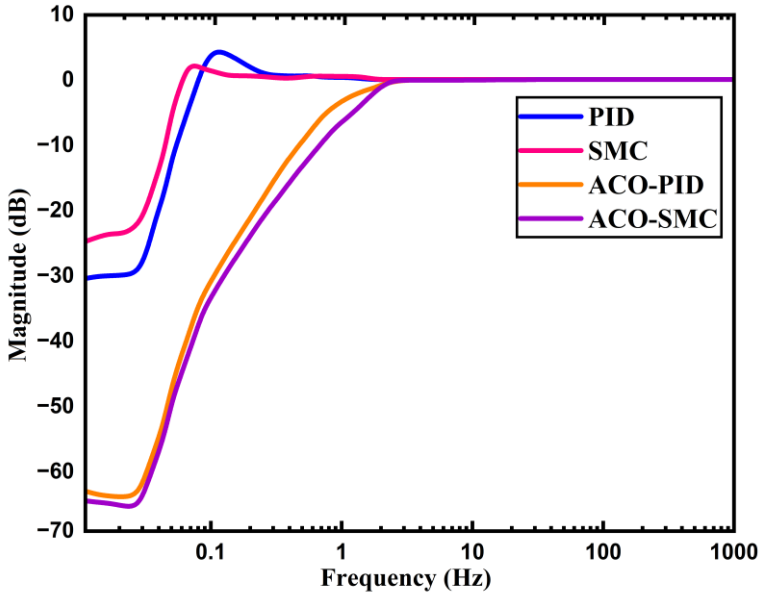


Figure 10. Frequency domain plots of the sensitivity function

The essential frequency-domain relationships for the sensitivity and complementary sensitivity functions can be expressed as follows:

$$S(s) = \frac{1}{1+L(s)} \tag{31}$$

$$T(s) = \frac{L(s)}{1+L(s)} \tag{32}$$

where, $L(s) = G(s)\,C(s)$ is the loop transfer function, $G(s)$ is the plant transfer function, and $C(s)$ is the controller transfer function.

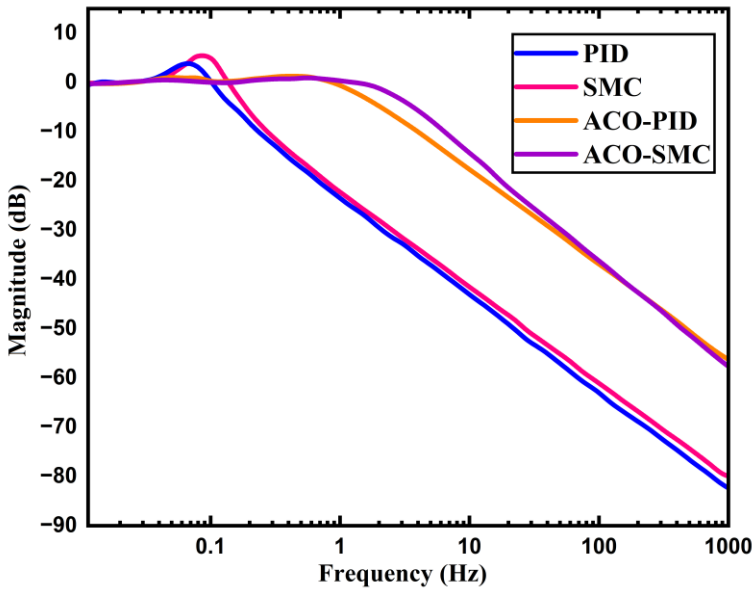


Figure 11. Frequency domain plots of the complementary sensitivity function

5. CONCLUSION

This study presented an improved ACO-SMC control strategy to enhance the speed control performance of an electric vehicle under standard driving cycles. The simulation results confirm that the proposed ACO-SMC controller achieved superior tracking performance, faster dynamic response, and significantly reduced tracking errors and chattering compared to PID, SMC, and ACO-PID controllers under both FTP-75 and US06 driving cycles. In addition, the proposed controller exhibits strong robustness against varying driving conditions, external disturbances, and system constraints, ensuring smooth and stable vehicle operation. However, this study is limited to simulation-based validation using a DC motor-driven electric vehicle model and does not include real-time experimental implementation. Furthermore, the inverter switching dynamics, thermal effects, and battery nonlinearities were not considered in detail. Future work will focus on the real-time implementation of the proposed controller on hardware platforms, extension to induction-motor-based electric vehicles, and experimental validation under practical operating conditions to further verify its effectiveness and applicability in real-world EV systems.

ACKNOWLEDGEMENT

The author would like to acknowledge the support, contribution, and sincere gratitude to the management of Vellore Institute of Technology, Vellore campus, Tamil Nadu, India.

REFERENCES

[1] Lee W J, Strbac G, Hu Z, Ding Z, Sarikprueck P, Teng F, and Kariniotakis G. (2020). Special issue on advanced approaches and applications for electric vehicle charging demand management. IEEE Transactions on Industry Applications 56(12): 5682–5683.

[2] Khooban M H, Niknam T, Blaabjerg F, and Dehghani M. (2016). Free chattering hybrid sliding mode control for a class of non-linear systems: electric vehicles as a case study. IET Science, Measurement & Technology 10(8): 776–785.

[3] Hermassi, M., Krim, S., Kraiem, Y., & Hajjaji, M. A. (2024). Adaptive neuro fuzzy technology to enhance PID performances within VCA for grid-connected wind system under nonlinear behaviors: FPGA hardware implementation. Computers and Electrical Engineering, 117, 109264.

[4] Arya Y. (2019). Impact of ultra-capacitor on automatic generation control of electric energy systems using an optimal FFOID controller. International Journal of Energy Research 43(14): 8765–8778.

[5] Patel V V. (2020). Ziegler–Nichols tuning method: understanding the PID controller. Resonance 25(12): 1385–1397.

[6] Hannan M A, Ali J A, Hossain Lipu M S, Mohamed A, Ker P J, Mahlia T M I, and Dong Z Y. (2020). Role of optimization algorithms based on fuzzy controllers in achieving induction motor performance enhancement. Nature Communications 11(1): 3792.

[7] Hou M, Zhao Y, and Ge X. (2017). Optimal scheduling of the plug-in electric vehicles aggregator energy and regulation services based on grid-to-vehicle. International Transactions on Electrical Energy Systems 27(12): e2364.

[8] Kashfi, R., Balochian, S., & Alishahi, M. (2024). Design of a optimal robust adaptive neural network-based fractional-order PID controller for H-bridge single-phase inverter. Applied Soft Computing, 166, 112142.

[9] Freitas, J. B. S., Marquezan, L., de Oliveira Evald, P. J. D., Peñaloza, E. A. G., & Cely, M. M. H. (2024). A fuzzy-based Predictive PID for DC motor speed control: JBS Freitas et al. International Journal of Dynamics and Control, 12(7), 2511-2521.

[10] Qu, S., Xu, W., Zhao, J., & Zhang, H. (2021). Design and implementation of a fast sliding-mode speed controller with disturbance compensation for SPMSM system. IEEE Transactions on Transportation Electrification, 7(4), 2611-2622.

[11] Chen, X., Zhao, J., Lu, Y., & Sheng, L. (2025). Research on sliding mode speed control method of permanent magnet drive system in shearer cutting section based on sliding mode observer under complex working conditions: X. Chen et al. International Journal of Dynamics and Control, 13(6), 234.

[12] Mousmi A, Abbou A, and El Houm Y. (2019). Real-time implementation of a novel hybrid fuzzy sliding mode control of a BLDC motor. International Journal of Power Electronics and Drive Systems 10(3): 1167–1177.

[13] Wu L, Liu J, Vazquez S, and Mazumder S K. (2021). Sliding mode control in power converters and drives: a review. IEEE/CAA Journal of Automatica Sinica 9(2): 392–406.

[14] Chan J W. (2022). Sliding mode control of brushless DC motor speed control. Malaysian Journal of Science and Advanced Technology 2(4): 188–193.

[15] Levant A. (2003). Higher-order sliding modes, differentiation, and output-feedback control. International Journal of Control 76(9): 924–941.

[16] Liu J, Zhao T, and Dian S. (2021). General type-2 fuzzy sliding mode control for motion balance adjusting of the power-line inspection robot. Soft Computing 25(2): 1033–1047.

[17] Flores Peña P, Luna M A, Ale Isaac M S, Ragab A R, Elmenshawy K, Martín Gómez D, and Molina M. (2022). A proposed system for multi-UAVs in remote sensing operations. Sensors 22(23): 9180.

[18] George M A, Kamat D V, and Kurian C P. (2021). Electronically tunable ACO-based fuzzy FOPID controller for effective speed control of an electric vehicle. IEEE Access 9(15): 73392–73412.

[19] Khooban M H, ShaSadeghi M, Niknam T, and Blaabjerg F. (2017). Analysis, control, and design of speed control of electric vehicles delayed model: multi-objective fuzzy fractional-order controller. IET Science, Measurement & Technology 11(2): 249–261.

[20] George, M. A., Kamat, D. V., & Kurian, C. P. (2024). Electric vehicle speed tracking control using an ANFIS-based fractional order PID controller. Journal of King Saud University-Engineering Sciences, 36(4), 256-264.

[21] Xu, M. (2022, November). Control of DC adjustable speed electric vehicle based on PSO-PID algorithm optimization research. In 2022 IEEE 5th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE) (pp. 616-621). IEEE.